Transport, Infrastructure and Growth: Evidence from Chinese Firms

Will Jianyu Lu*

Central Bank of Chile

January 2021

Abstract

Transportation infrastructure is a way for governments to achieve development goals by connecting peripheral areas to urban centers. Yet understanding the gains from transportation infrastructure is difficult because the placement and timing of transportation projects are not random. This paper studies the effects of China's National Trunk Highway System (NTHS) on firms in the peripheral regions. I construct a new data set of completion times of every segment of the highway system and combine it with a comprehensive data set of Chinese manufacturing firms. To address the endogeneity of placement and timing of the highway construction, I develop a novel time-varying instrumental variable based on a civil engineering model of least-cost construction and local completion rates. The IV results show that firms connected in the peripheral regions experience faster growth in output and sales. To understand the mechanism, I study firms' input choices. Highway connection increases the firms' growth in intermediate inputs and capital, while the growth of labor declines. This is consistent with a decline in the prices of capital and intermediate inputs due to access to outside markets. Using a time-varying instrument is crucial for demonstrating these effects. When using methods similar to previous studies, I find null or opposite results.

* *Contact Info*: Email: wlu@bcentral.cl. The views expressed herein are those of the author and not necessarily those of the Central Bank of Chile, or its Board members.

1 Introduction

Access to transportation infrastructure is seen as a crucial element in a nation's growth and development. It reduces the cost of trade and migration, equalizes prices, and facilitates the spread of ideas and technology. Thus, in recent decades, large investments have been devoted to building roads. For example, the World Bank allocates about 12% of its total lending to upgrading road networks in developing countries. Moreover, China's National Trunk Highway System (NTHS) is estimated to have a construction cost of 2 trillion RMB (290 billion US dollars).

There are many possible gains from improved transportation infrastructure. For instance, it can reduce trade cost and inter-regional price gaps (Donaldson, 2018) and lower firms' inventory holdings (Datta, 2012), or decrease the risk of loss of market access (Brooks and Donovan, 2020). However, clear identification of the causal effects of an infrastructure project such as a road network remains a challenge. This is due to the potential endogeneity of the placement of the roads.[1] The existing literature addresses this issue by using instruments based on counterfactual networks. However, these are static instruments that cannot capture the time-varying aspect of large road networks rollout. These results are mixed with some papers finding positive results and others finding negative results from transportation infrastructure.

I study China's National Trunk Highway System (NTHS), the largest highway construction project in world history. Its objective is to connect provincial capitals and cities with

[1]For example, a highway planner may channel road construction resources towards locations growing unusually fast (if there are complementaries between transportation and growth) or slow (if there are political economy reasons to support struggling areas). Either of these cases would bias the estimated effects of roads.

an urban population above 500,000. Given its massive total length (over 110,000 km), a secondary effect is for some non-targeted peripheral locations between the large metropolitan centers to also become connected. I focus on the firms in these peripheral regions, away from the metropolitan centers.[2]

In this paper, I combine detailed data on the timing of road construction with a comprehensive dataset of Chinese firms to measure the effect of road construction on firm outcomes. I develop a novel instrumental variable strategy to identify the effects of road construction and I show why previous papers have found negative or null results. Using my IV strategy, I find that connecting firms in the periphery leads to higher output and sales growth as well as increased labor productivity. When exploring the mechanism, I find that firms' usage of intermediate inputs and capital grow significantly, but the growth rate of labor employed declines. The differences among the growth of input suggest that the prices of intermediate inputs and capital have decreased, which is consistent with lower trade cost due to better roads. The key econometric force behind these results is a time-varying instrument. To show the importance of the dynamic instrument, I remove the time variation of the highway connection. This flips the results, bringing them in line with other work using static instruments.[3]

I construct a time series of China's National Trunk Highway System by locating and dating completion dates of 470 segments belonging to the system. To ensure accuracy, we

[2]The large cities are excluded because they are the NTHS connection targets and the identification strategy used in this paper is not suited to evaluate them.

[3]My results differ from previous papers that find negative null results. For instance, Faber (2014) finds that China's NTHS lead to a *reduction* in industrial and total output growth among the connected peripheral regions relative to non-connected ones. However, his approach uses a ten-year growth rate difference, static instrument, and county-level data, rather than annual growth rates, a dynamic instrument, and firm-level data as used in this paper. I examined a static variants of my specification and found similar results.

relied on multiple sources. To determine dates, I rely on official newsletters from the highway authority, State Council Announcement, and local newspapers. Then I combine this with a comprehensive dataset on every medium and large manufacturing firm in China. The firm-level data includes all state-owned firms as well as non-state-owned firms whose sales are more than five million RMB (about 645,000 US dollars in 2007) per year. The variables that I explore include output, sales, inputs, and other measures.

Even though peripheral regions are not explicitly targeted by the NTHS planning process, there might still be endogeneity involved with the highway placement. To address these concerns, I develop a new instrumental variable approach to address the endogenous placement and timing of the highways. My instrument has two components, one deals with location and the other with timing. To address the unobserved biases in the placement of the NTHS, I construct a counterfactual least-cost path network using standard engineering algorithms similar to those utilized in Faber (2014). To add time variation to my instrument, I rely on the yearly completion rate of actual highways at the province-level. I compute the total length of highways built in a province in a particular year as a percentage of the total length of highways planned in that province. Then I use the computed percentage as the likelihood that locations on the counterfactual network become connected. My final instrument is the combination of the two components.

The empirical approach of this paper is related to recent studies on transport infrastructure. Banerjee et al. (2020) use straight-line connections between pairs of historical cities and "treaty ports" to predict the construction of railway lines in the late 19th and early 20th century China. Baum-Snow et al. (2017) use the configuration of urban transportation infrastructure in 1962 as instruments for more recent transportation infrastructure to achieve exogenous variation in the identification. Faber (2014), whose instrument resembles closely the geospatial component of mine, constructs hypothetical networks based on straight-lines and least-cost path using geographical characteristics. All the strategies mentioned are static; they do not vary with time and cannot capture the time-series variations of the transportation infrastructure involved. I deploy a dynamic instrument that is capable of estimating the precise time-varying effect of improved road networks, and I show that the dynamic aspect of the instrument is crucial for estimates. [4]

This research is related to a growing literature on the study of the regional effects of improved transportation networks. Michaels (2008) finds that the construction of the interstate highway system has led to changes in skill premia across counties in the U.S. Banerjee et al. (2020) determine that proximity to transportation networks has a moderate positive causal effect on per capita GDP levels across sectors in China. Asturias et al. (2018) quantify the effects of the Golden Quadrilateral in India and decompose the welfare effects into pro-competitive and allocative efficiency. Relative to the existing studies, this paper uses detailed microdata to precisely estimate the causal effects of better roads on firms. With panel data, I can provide more insights on the channels and mechanisms at work. My findings that roads increase capital intensity may relate to the skill premium if capital-skill complementarity is pervasive (Krusell et al., 2000).

The next section presents the background on NTHS and the datasets used in this paper. Section 3 introduces the empirical approach. Section 4 presents the main estimation results.

[4] Yang (2018) also constructs a time-varying least-cost path spanning-tree network as an instrument for actual highway connections to evaluate China's highway expansion. For time variation, his instrument relies on national highway construction budget, whereas my instrument is based on province-level completion rate. Note that provincial governments finance about 70% of the construction of NTHS within each province. See Section 3.2 for additional information.

Section 5 presents alternative specifications to reconcile with the findings of the existing literature. Section 6 concludes.

2 Background and Data

In this section, I describe the planning and characteristics of China's NTHS, and the dataset used and constructed for the purpose of studying the impact of largest expressway infrastructure in the world. This section is structured as follows. Subsection 2.1 describes the policy background. Subsection 2.2 describes firm-level data. Subsection 2.3 describes trade data and Subsection 2.4 presents the geospatial data.

2.1 Policy Background

China's NTHS was originally planned in 1992 by the Chinese State Council, in order to connect all major cities in one single network. The initial network was known as "5-7" network, which refers to five vertical (north-south) and seven horizontal (east-west) routes. It had the objective of connecting all provincial capitals and cities with an urban registered population above 500,000 with the nation's capital Beijing. The network had a total length of 35,000 km and was scheduled to be completed in 15 years. However, most of the routes were completed ahead of schedule. Therefore, in late 2004 the central government issued a revision to expanded the original plan. The revised construction plan was known as the "7-9-18" system because there are 7 radial expressways departing from the national capital of Beijing, 9 north-south expressways and 18 east-west expressways. Figure 1 displays the network in 2013.

The NTHS is a highly improved system of expressways compared to the previously existing road networks. The NTHS routes are limited access toll roads with at least four lanes (sometimes six lanes or even eight lanes). These expressways are superior in road condition and driving speed relative to the pre-existing national and provincial highways. The expressways in NTHS have a maximum speed limit of 120km/h and a minimum of 70km/h, and penalties for driving below or in excess of the prescribed speed limits are strictly enforced. In contrast, other highways and roads in China usually have a maximum speed limit of 70km/h. Therefore, the expressways of NTHS significantly reduce the travel time. For example, when the MeiHe Highway opened in 2006, the driving time between Guangzhou and Meizhou, a within province drive, decreased from 6 hours to 4 hours.

The estimated cost for the "7-9-18" system is 2 trillion RMB (about 290 billion US dollars in 2018). The central government only provides limited funding for the construction of NTHS, only about 10% of the total cost, with the rest mainly raised by the provinces. The provincial and local governments use their budget from vehicle license fees and purchase tax to cover a minor portion of the national highway system construction cost and finance most expenses by borrowing against future tolls.[5]

The construction speed of the NTHS has been unprecedented: the network grew from less than 5,000km in 1998 to over 60,000km in 2009. Figure 1 depicts the segments completed and planned until 2013. A large number of firms in the peripheral regions became connected over the years. In 1999, only 20% of the firms in the peripheral regions were connected by NTHS, that number grew to 46% in 2004 and 72% in 2007. Both the variation in the

[5] This fact will be explored in the construction of the time-varying component of my instrument. See Section 3.2 for details.

location of placement of highways as well as their completion dates are used to explore the effects of improved road network on firms. Figures for the progress of highway construction can be found in Appendix A.1.

2.2 Firm-level Data

The data for Chinese firms are from the Annual Survey of Industrial Production from 1999 to 2007 conducted by the National Bureau of Statistics of China (NBS). It has been quite commonly used in development economics studies such as Cai and Liu (2009), Hsieh and Klenow (2009) and Khandelwal et al. (2013). The dataset includes all state-owned firms as well as non-state-owned firms whose sales are more than five million RMB (about 720,000 US dollars in 2018) per year, spanning 37 two-digit industries and 31 provinces or province-equivalent municipal cities. In this study, I drop utility and mining industries because they are likely less affected by the highway system.[6] The raw data contains over 100,000 firms in 1999 and it grows to over 300,000 in 2007.

Firm-level data for the NBS have several distinct advantages over GDP data estimated and reported by local governments. NBS's data is superior in accuracy, breadth, and detail. In addition, the panel aspect of firm-level data allows for fixed effects, as well as the notion of entry and exit.[7] By using the Annual Survey of Industrial Production, I can achieve much more precise estimates for the impact of the NTHS on the peripheral regions. This

[6] Both utility and mining industries do not mainly rely on roads to transport their products. Utilities have their own delivery infrastructure, and mining production uses primary railroads for transportation.

[7] Cai and Liu (2009) claims that "*The information reported to the NBS should be quite reliable, because the NBS has implemented standard procedures in calculating the national income account since 1995 and has strict double checking procedures for 'above-scale' firms. Moreover, firms do not have clear incentives to misreport their information to the NBS, because such information cannot be used against them by other government agencies such as the tax authorities. Misreporting of statistical data was commonly suspected for some time in China, the most notorious was local GDP data provided by local governments.*"

is a notable advantage over existing literature on the impact of the NTHS that uses the GDP data at county-level. As opposed to local aggregate level data, firm-level data is more reliable and detailed. Variables of interest in this survey include industry, employment, age, ownership, wage payment, value-added, intermediate inputs, export revenues, inventory, wage payment, debt, profits, investment, fixed assets, and sales.[8] I use the six-digit postcode to locate the firm's county and then use the county location to link to my geospatial dataset described in Section 2.4.

To study firm dynamics for this large transportation improvement, keeping track of firms over time is essential. To guarantee that the panel information of firms is accurate, I use different combinations of firms' identification number, postcode, name (both key information and full name), main representative to match firms over the years. As a result, I have identified over 535,000 unique firms out of 1.8 million observations over nine years. Of these, I can construct a balanced panel for over 33,000 firms over nine years. The panel data is used to study the dynamics firms' outcomes resulting from connection to the NTHS.

2.3 Trade Data from Customs Transactions

The trade data are collected from import and export shipping manifests by the customs agency in China. The dataset ranges from 2002 to 2006 and gives detailed information for each trade shipment. In most cases, each entry includes the name of the importer or exporter, the month of declaration, source or destination country, quantity, unit price, and value of the good, along with detailed information such as the 10-digit Harmonized System

[8]Descriptive statistics can be found in the Appendix A.1. Since I observe firm age, I can also study firm entry.

(HS) classification.

I use this dataset to understand if firms' access to imported inputs, particularly intermediate inputs and capital goods, changes with NTHS connection. Based on Broad Economic Categories (BEC) Revision 4, I focus on the import transactions involving HS codes belonging to intermediate inputs (BEC categories 111, 121, 21, 22, 31, 322, 42 and 53) and capital goods (BEC categories 41 and 521). The total amount of imported inputs is the sum of the two types of goods.[9] In my analysis, I restrict to my trade data to transactions involving imports of firm inputs. I also consider BEC category 4*, "Capital goods (except transport equipment), and parts and accessories", for the import of machinery.

2.4 Geospatial Data

This section describes the construction of the time-series of the NTHS and the use of geographic information system (GIS) in this project. In order to better study the impact of the highway system and differentiate the short and long run effect, it is necessary to have precise dates that each highway segment was completed. The existing road data are usually from digitalized maps that are constructed infrequently and between irregular time intervals. In addition, digitalized maps are sometimes inaccurate.[10] Therefore, I have carefully constructed with high precision every segment of the NTHS and date the completion year of each individual segment.

First, I locate the 34 highways that constitute the NTHS that was completed in 2013. Figure 1 displays the network in 2013, which has a total extension of 97,355 km (60,494

[9] Using equivalence from United Nations (1989), I obtain the relevant HS codes that involve capital goods.

[10] Also I find some discrepancy and inaccuracy in the Chinese roads used by Faber (2014). For example, the maps in his paper show that there is a NTHS expressway connecting the province of Qinghai and Tibet in 2007, however, as of 2015, this portion is yet to be constructed.

mi). Figure 2 shows the progress of the NTHS in 1999, 2003 and 2007. In order to precisely located each highway, I verify the roads with OpenStreetMaps, Baidu Map, and Chinese road atlases. In order to date each highway, I split the whole system into 470 segments. My main source of data to date the highways are the official newsletters and local newspapers. I also seldom use Google Earth Historical Imagery. For most the segments, I obtained the completing year, month and day, but in order to utilize the time series with the firm-level data, I only focus on the year. Since the industrial surveys contain calendar year information of the firms, we set all the completion year to the following year. For example, roads that have completion date between January 1, 2003 and December 30, 2003 will show as completed in 2004.

Other geographic location data such as county boundaries and postcodes are obtained from the GADM database of Global Administrative Areas. China can be divided into 33 provinces or province equivalent units, 347 prefectures (known as prefectural-cities) and 2341 counties. I will use the county as the geographical unit in this paper. Land cover and elevation data used to construct the instrumental variable were obtained from the United States Geological Survey of Digital Chart of the World project.

3 Empirical Approach

This section introduces the empirical approach to estimate the effects of the NTHS on firms that it connects. Subsection 3.1 describes the baseline OLS estimation. Subsection 3.2 describes the instrumental variable that is used to address the endogenous placement and timing of the highways. Subsection 3.3 proposes a specification to study the heterogeneous effects of the NTHS connection on firms with different characteristics.

3.1 Baseline Estimation

First, I classify counties into three types depending on their location relative to the NTHS and to the targeted cities: (i) *targeted*, (ii) *connected peripheral*, and (iii) *non-connected peripheral*. Targeted counties are those counties within 50 km radius from the center of a targeted city. There are 54 large metropolitan cities and provincial capitals that the NTHS intended to connect. I exclude the *targeted* firms in the estimation because NTHS is not the main factor affecting their performance. A firm that is not *targeted* is a *peripheral* firm. If a *peripheral* firm is in a county within a 10km distance from the NTHS, it is a *connected peripheral* firm. Otherwise, it is a *non-connected peripheral* firm.

The baseline regression specification (1) explores difference across time (growth), difference across connected firms and not, and difference in growth before and after connection.

$$\ln\left(y_{i,t}\right) - \ln\left(y_{i,t-1}\right) = \alpha_i + \gamma_t + \beta_1 Connect_{i,t} + \eta X_{i,t} + \epsilon_{i,t} \tag{1}$$

where $y_{i,t}$ is an outcome of interest of firm i in year t, α_i is a firm fixed effect, γ_t is a time fixed effect, $Connect_{i,t}$ is a dummy that indicates whether firm i is connected in year t. $X_{i,t}$ is a vector of firm control variables. The firm outcomes that I examine are output, total sales, profit, value-added, intermediate inputs, capital, labor, and wages. The firm characteristics that I control for are log of total assets, State-Owned Enterprise status and its interaction with connectivity.

3.2 Instrumental Variable Approach

Estimating the baseline specification by OLS requires a potentially problematic assumption that counties connected by the NTHS were done so in random fashion. This assumption might be plausible if the NTHS was constructed strictly following the original plan. However, provincial governments are responsible for raising credit and paying for the construction of the highway system. Sometimes local planners may choose to place the road in areas where they expect growth to be faster in the future. Therefore, both placement and timing of the highway are potentially endogenous to the characteristics of locations that were connected. To address this issue, I create an instrument in two steps: first, I create a least-cost path that solves the endogeneity in the location of the highway placement, and then I use the annual completion rate of actual highways at the province-level to add time variation to my instrument. In principle, fixed effects might account for much of this heterogeneity, but not if locations had differed in time-varying ways. The timing of road construction is, therefore, another potential issue.

In the first step, I construct a hypothetical least-cost path spanning tree network (Figure 2) to proxy for the actual route placement. This counterfactual network is based on the global minimum construction cost to connect all the targeted cities in a single network. In other words, they are the routes that the Chinese State Council would have chosen for the planned policy if construction cost were the only driving factor. The least-cost path spanning tree network is the resu‘lt by combining two methods. The first method is Dijkstra (1959)'s optimal route algorithm to compute the most economical path between each pair of targeted cities relying on remote sensing data on land cover and elevation. Next, I apply Kruskal

(1956)'s minimum spanning tree algorithm on the previously obtained least-cost bilateral routes to compute a network with the least bilateral connections but still connecting every targeted city.

Second, to account for the endogenous timing of the NTHS construction, I construct an instrument based on the yearly highway construction in each province. I calculate the actual length of highways constructed in each province every year as a percentage of the total length of highways in that province. Then I use this percentage as the probability that locations on the hypothetical network become connected. Since provincial governments are in charge of the financing of the highway construction in their province, it is reasonable to assume that their budget is fixed. Therefore, more of the actual highway they build is positively correlated with how much they would have built for the counterfactual highway.

So the IV is constructed by two variables: the connection to the engineering minimum span tree lowest cost and the percentage of the actual highway built in a province as a proportion of the total planned highway in that province.

$$ConnectIV_{c,t} = ConnectIV_c * PercentBuilt_{(p,t)} \tag{2}$$

where $ConnectIV_c$ indicates the connection status of county c to the contra-factual highway and $PercentBuilt_{(p,t)}$ is the percentage of the actual highway built at time t in province j that county i belongs. Figure 3 shows the implementation of the counterfactual highway over time.

Table 1 presents the first stage for my instrumental variable at the county level. The instrument is a strongly significant predictor of the actual NTHS placement and timing.

There might be concerns that province specific growth patterns may affect the results due to the component $PercentBuilt_{(p,t)}$, so robustness checks using province linear trend, province-time fixed effects, and percentage built as control are included in Appendix A.2.

3.3 Firm Heterogeneity Analysis

The effect of becoming better connected to other markets may not be equal for firms of different types. In this subsection, I design a regression strategy to identify the heterogeneous effects that connection to the NTHS may have on firms of distinct characteristics. I add controls for firm characteristics, which may differentially influence the NTHS impact, the specification in equation (1).

$$\begin{aligned} \ln(y_{i,t}) - \ln(y_{i,t-1}) = \alpha_i + \gamma_t + \beta_1 Connect_{i,t} + \beta_2 z_{i,t} \\ + \beta_3 z_{i,t} * Connect_{i,t} + \eta X_{i,t} + \epsilon_{i,t} \end{aligned} \tag{3}$$

where the new variable $z_{i,t}$ is the characteristic to control for the heterogeneous effect. β_3 will capture the heterogeneous effect. The dimensions I consider are firm age, size (defined by total assets), initial export status, total employment and debt to asset ratio.

4 Main Estimation Results

This section reports OLS and IV estimation results of specification (1) for a number of different firm-level outcomes. Subsection 4.1 presents the results on growth for output, sales, profit, value-added and value-added per worker for all firms in the peripheral regions. Subsection 4.2 presents the same outcomes as the previous subsection but restrict the sample to only firms that existed before the highway connection. Subsection 4.3 explores the channels of growth via firm inputs.

4.1 Firm Growth

This subsection reports the estimation results of specification (1) for a set of firm performance outcomes. Table 2 presents OLS and IV results of regressing change of growth rate on the binary NTHS treatment variable and including all the firms observed in the data that are in the *peripheral* counties. There are two main takeaways from Table 2 –connection has positive impacts on firm growth, but the instrument is necessary to discern this result.

The OLS estimates of the NTHS connection effect are positive but not statistically significant, and the IV results are not only larger in magnitude but also statistically significant. This suggests that the planner may have had additional plans to develop particular *peripheral* counties where growth was expected to be none or slow, so they may have additionally targeted to connect those places with the NTHS.[11] There is more evidence on slow-growing peripheral counties being targeted by the NTHS in the subsection on pre-existing firms.

The second clear pattern is that the inclusion of controls such as size and state ownership attenuate the effects of NTHS. Size is actually an important influencer in growth: larger firms in China tend to grow faster. State ownership status is a negative driver of growth; State Owned Enterprises are likely to decline over time.

The two patterns mentioned above hold for all dependent variables mentioned in Table 2

[11]The NTHS planner may have favored certain route placement in the *peripheral* region for reasons such as corruption, personal goals, and other socio-economic incentives. Alder and Kondo (2020) find that birthplaces of the top officials who were in power during the network's planning are more likely to be closer to the NTHS.

with the exception of profit, which does not show any statistically significant result. Output is the total value of the final product produced, sales are the total value of revenue from the final product sold and profit is the financial book balance. Value-added in my dataset is simply the difference between the value of output and the value of intermediate good consumption. The large magnitude of the IV Connect on value-added will be addressed in the subsection 4.3. In addition, robustness checks using province linear trend, province-time fixed effects and percentage built as control can be found in Appendix A.2.

4.2 Pre-existing Firms

There are many heterogeneities among the firms in *connected* counties. One of the main characteristics that may affect the causal effect of the NTHS connection is whether a firm started its activity before or after its location was connected by the expressway system. In this paper, an pre-existing firm is one that opens in a location that is not yet connected by the NTHS. On the other hand, a new entrant is a firm that locates itself in a county that is already connected by the expressway system. Due to their starting time, new entrants have the ability to choose their location such that the observed NTHS connection effect may not be entirely causal.

This subsection reports the estimation results of specification (1) for a set of firm performance outcomes and restricts the data to only include pre-existing firms. Similar patterns as Table 2 are observed in Table 3. Although OLS and IV estimates are both statistically significant, OLS is negative while IV is positive. These results show that NTHS connect is negatively correlated with *peripheral* counties growth, suggesting that possibly the planner has chosen locations that are expected to grow significantly slower. However, the expressway connection still leads to higher firm growth.

By my definition, pre-existing firms are those that did not choose their location to be connected by the NTHS. Therefore, it is reasonable for the magnitude results on this subset of the data to be lower than the full dataset, which includes the new entrants. However, we still observe that IV connection has a positive and statistically significant effect on firm growth rate of output, sales, and value Added. To look further in the channels of the increasing growth, we will turn to the inputs.

4.3 Mechanism through Inputs

This subsection reports estimation results of specification (1) for firm inputs to explore how the NTHS affect firm performance. Table 4 presents OLS and IV results on intermediates inputs, employment and total assets of the firm when restricting the dataset to only include pre-existing firms. I will focus on the IV results because OLS results are possibly biased by the roads placement. In line with previous firm performance results, firms on the NTHS uses more intermediate inputs and capital in their production. Surprisingly, Table 4 also shows that these growing firms are employing fewer workers. At first sight, it may be hard to reconcile the declining growth of employment with a higher growth rate of output. There are two channels that can support this finding: the relative price of inputs and access to labor-saving imported inputs.

The first channel, relative price of inputs, can be explored via the cost minimization of any standard production function. For expositional purposes, I assume Cobb-Douglas production function.

$$\min_{l,k} \quad wl + rk \quad \text{s.t.} \quad y = zl^{\alpha}k^{\beta} \tag{4}$$

The optimality condition implies that the demand for labor l can be written as a function of factor prices and the quantity produced.

$$l(w, r, y) = \left(\frac{r}{w}\frac{\alpha}{\beta}\right)^{\frac{\beta}{\alpha+\beta}} y^{\frac{1}{\alpha+\beta}} \tag{5}$$

From the results in Table 4, we clearly see that demand for labor l shrinks and output y rises. Therefore for equation (5) to hold, it must be the case that the rental rate of capital to wages ratio decreases ($\frac{r}{w} \downarrow$). The rental rate of capital is not observed in the data, but the average wage can be computed. Table 5 displays the effects of NTHS on the growth of average wages. The wage in connected firms does not change significantly, so for $\frac{r}{w}$ to fall, it must be that the rental rate of capital has decreased.

This result is consistent with trade theory, lower trade cost due to better roads decreases the prices of intermediate inputs and capital. As a result, firms become more capital-intensive and their labor share falls. Considered together with the unchanged wages, it is suggestive that workers are not reaping the benefits from firm's better performance. Another puzzling finding is that output growth increases despite the decline of labor growth, so it could be the case that connected firms are getting better technology. When firms are connected by the highway, they are also being connected to ports that give them access to the international market of goods. So connected firms gain more access to better inputs from abroad.

Next, I explore the second channel by looking at the use of foreign inputs. First, I aggregate the trade data to the county level to examine the impact of NTHS connection on the value of imported inputs. Counties that do not have firms importing inputs from abroad are assigned a value of zero.[12] Table 6 presents IV results on Imported Inputs for counties in the peripheral regions. Although it shows that there is a generalized pattern of connected counties importing more foreign inputs, only the coefficients on imported capital goods are statistically significant.

To further examine the previous results, I look at the imports of capital in particular. Researchers have found that imported capital goods are major carriers of R&D spillover (Xu and Wang, 1999), translate into productivity gains (Eaton and Kortum, 2001) and increase significantly the per capita income growth rates across developing countries (Lee, 1995). In most developing countries, bad roads are great impediments to import large and advanced machinery. With the introduction of NTHS, access to imported capital goods, in particular, for connected firms in the peripheral regions may have significantly improved. Table 7 presents the effect of NTHS on the growth rate of imported capital goods on firms in the peripheral regions.[13] The connection indicator is positive and statistically significant suggesting that the import of machinery and other manufactured goods used by industries increased nearly 60% for the connected firms relative to the non-connected firms.[14] China's National Trunk Highway System seems to have allowed connected firms to import more and better technology from abroad and thus allowed them to become much more productive.[15]

[12] To compute the growth rate of imports involving zero values, I use three methodologies by using differences in inverse hyperbolic sines, differences in logarithms and setting the zero values to 1, and computing the mean growth rates.

[13] See Section 2.3 to learn how this dataset is constructed.

[14] Note that these firm-level estimations only capture the intensive margin of imports (conditional on importing firms). After I link the firm-level data with the trade data, both intensive and extensive margins will be explored.

[15] This is also consistent with studies in the growth and trade literature, such as(Sala i Martin et al., 2004),

5 Dynamic vs. Static Specifications

Most existing empirical literature including this paper finds that new transport infrastructure are beneficial to locations that become better connected (Banerjee et al., 2020; Datta, 2012; Ghani et al., 2015). However, Faber (2014) who studies the same infrastructure project as this paper finds that China's NTHS has led to a reduction in GDP growth among connected peripheral regions. His approach uses a static instrument variable and county-level data, which are different from the time-varying instrument and firm-level data used in my approach. Therefore, I examine alternative specifications to find the source of differences between my results and those of Faber (2014).

The first alternative specification aggregates the firm-level data to county-level. Specification (6) is a modified version of the baseline regression for aggregate level data.

$$\ln(y_{c,t}) - \ln(y_{c,t-1}) = \alpha_c + \gamma_t + \beta_1 Connect_{c,t} + \eta X_{c,t} + \epsilon_{c,t} \tag{6}$$

where $y_{c,t}$ is an outcome of interest of county c in year t, α_c is a county fixed effect, γ_t is a time fixed effect, $Connect_{c,t}$ is a dummy that indicates whether county c is connected in year t. $X_{c,t}$ is a vector of firm control variables. The county outcomes that I examine are output, total sales, and value-added. The county characteristic that I control for is the log of total industrial assets. Table 8 shows OLS and IV results of the specification (6) county-level regression. The pattern remains the same as firm-level results, the sign of the IV estimates remains positive but they are no longer statistically significant. This evidence suggests that

which finds that a low price of capital goods is positively related to growth. Better roads lowers the price of acquiring capital goods.

aggregating the firm-level data to county-level causes a loss in precision but the positive impact of NTHS does not change.

Then, I introduce a second alternative specification by taking a long-difference in growth instead of yearly growth rate. Still using the firm-level data, this is an approximation to Faber (2014)'s static approach. I use a nine-year difference, which is quite comparable to Faber's ten-year difference. Since I use a nine-year growth rate, the sample size is significantly reduced. Specification (7) is a modified version of the baseline regression to study long-differences.

$$\ln(y_{i,2007}) - \ln(y_{i,1999}) = \alpha_p + \beta_1 Connect_{i,2007} + \eta X_i + \epsilon_i \tag{7}$$

where $y_{i,t}$ is an outcome of interest of firm i in year t, α_p is a province fixed effect, $Connect_{i,2007}$ is a dummy that indicates whether firm i is connected in the year 2007. X_i is a vector of firm control variables. The firm outcomes that I examine are aggregate output, total sales, profit, and value-added. The firm characteristics that I control for are log distance to the nearest targeted city, the log of total assets, State-Owned Enterprise status and its interaction with connectivity. Table 9 presents OLS and IV results of the specification (7). With a nine-year growth rate as the outcome, the estimated effect of NTHS on firms is negative, opposite to the main findings of this paper. It is clear that the main source of difference between Faber (2014) and this paper originates from the fact that he uses long-difference of growth rate whereas I use annual growth rate difference.

The third alternative specification most closely resembles Faber (2014)'s approach. It aggregates firm-level data into county-level and uses a nine-year growth rate difference.[16]

[16] A minor difference is that Faber (2014) uses a ten-year growth rate difference between 1997 and 2006 whereas I use a nine-year growth rate difference between 1999 and 2007.

Specification (8) combines changes from the previous two alternative specifications.

$$\ln(y_{c,2007}) - \ln(y_{c,1999}) = \alpha_p + \beta_1 Connect_{c,2007} + \eta X_c + \epsilon_c \tag{8}$$

where $y_{c,t}$ is an outcome of interest of county c in year t, α_p is a province fixed effect, $Connect_{c,2007}$ is a dummy that indicates whether county c is connected in the year 2007. X_c is a vector of county control variables. The county outcomes that I examine are output, total sales, and value-added. The county characteristic that I control for is the log of total industrial assets. Table 10 shows OLS and IV results of the specification (8) county-level regression. The estimation results show similar a pattern to the second alternative specification.

5.1 Argument for Time Series Data

High-frequency time series is superior to a long-difference dataset, particularly in fast evolving economic settings such as China. Besides, China's NTHS has a progressive rollout. In 1999 only 15.3% of counties were connected whereas by 2007 50.2% of counties had access to the NTHS. With a significant portion of the counties (about 27%) getting the NTHS connection in the last three years of the sample, it is very likely that a nine-year long-difference is unable to correctly identify the effects of the improved transportation network. This issue is even greater concern if the counties connected near the end of the sample period have lower growth than the non-connected counties. In such a case, pooling the firms connected for many years with those connected for a few years can actually lead to finding negative effects.

6 Conclusion

Regional differences are particularly severe in developing countries. Large-scale transportation infrastructure such as China's NTHS can integrate markets in peripheral regions with large metropolitan centers by lowering the trade cost between them. The common intuition is that a better road network should have beneficial impact on regions that it connects. However, Faber (2014) exploits China's NTHS as a source of exogenous variation and find that this large transport infrastructure has led to a reduction in industrial and total output growth among connected peripheral regions.

This paper uses detailed firm-level data, highway construction data and a time-varying instrument to study the effects of China's NTHS on firms in the peripheral regions. The empirical results show that the NTHS connection has led to faster growth in output and sales as well as labor productivity. Exploring the channels via firms' inputs, I find that connected firms also have higher growth in intermediate inputs and capital but their labor growth rate declines. Since average wages do not change, the result on inputs suggests that the price of capital and intermediates may have declined, which is consistent with lower trade cost. Using trade data, I find that firms in the connected peripheral regions import more foreign capital goods. This finding is suggestive of more implementation of labor-saving technologies.

The results of this paper rely on a dynamic instrument. Removing the time variation of the highway connection reconciles with the results of earlier work that uses static instruments. Large transportation infrastructure such as the NTHS takes several years to complete, so locations on the network get connected at different times. Ignoring the timing of the highway rollout pools together the outcome of regions connected for many years with the ones connected for only a few years, thus it may not account for the actual impact of improved road network.

References

Alder, Simon and Illenin Kondo, "Political Distortions and Infrastructure Networks in China: A Quantitative Spatial Equilibrium Analysis," *Working Paper*, 2020.

Asturias, Jose, Manuel Garca-Santana, and Roberto Ramos, "Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, 11 2018, *17* (6), 1881–1940.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the road: Access to transportation infrastructure and economic growth in China," *Journal of Development Economics*, 2020, *145*, 102442.

Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang, "Roads, Railroads, and Decentralization of Chinese Cities," *The Review of Economics and Statistics*, 2017, *99* (3), 435–448.

Brooks, Wyatt and Kevin Donovan, "Eliminating Uncertainty in Market Access: The Impact of New Bridges in Rural Nicaragua," *Econometrica*, 2020, *88* (5), 1965–1997.

Cai, Hongbin and Qiao Liu, "Competition and Corporate Tax Avoidance: Evidence from Chinese Industrial Firms," *The Economic Journal*, 2009, *119* (537), 764–795.

Datta, Saugato, "The impact of improved highways on Indian firms," *Journal of Development Economics*, 2012, *99* (1), 46 – 57.

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

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, April 2018, *108* (4-5), 899–934.

Eaton, Jonathan and Samuel Kortum, "Trade in capital goods," *European Economic Review*, 2001, *45* (7), 1195 – 1235. International Seminar On Macroeconomics.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *The Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Ghani, Ejaz, Arti Grover Goswami, and William R. Kerr, "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing," *The Economic Journal*, 2015, pp. n/a–n/a.

Hsieh, Chang-Tai and Peter Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009, *124*, 1403–1448.

i Martin, Xavier Sala, Gernot Doppelhofer, and Ronald I. Miller, "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *The American Economic Review*, 2004, *94* (4), 813–835.

Khandelwal, Amit K., Peter K. Schott, and Shang-Jin Wei, "Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters," *American Economic Review*, 2013, *103* (6), 2169–95.

Krusell, Per, Lee E. Ohanian, Jos-Vctor Ros-Rull, and Giovanni L. Violante, "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica*, 2000, *68* (5), 1029–1053.

Kruskal, Joseph B, "On The Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem," *Proceedings of the American Mathematical society*, 1956, *7* (1), 48–50.

Lee, Jong-Wha, "Capital Goods Imports and Long-Run Growth," *Journal of Development Economics*, 1995, *48* (1), 91 – 110.

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

United Nations, "Classification by Broad Economic Categories (BEC) Revision 4," *Statistical Papers*, 1989, *Series M* (53).

Xu, Bin and Jianmao Wang, "Capital Goods Trade and R&D Spillovers in the OECD," *The Canadian Journal of Economics*, 1999, *32* (5), 1258–1274.

Yang, Yang, "Transport Infrastructure, City Productivity Growth and Sectoral Reallocation: Evidence from China," *IMF Working Paper*, 2018.

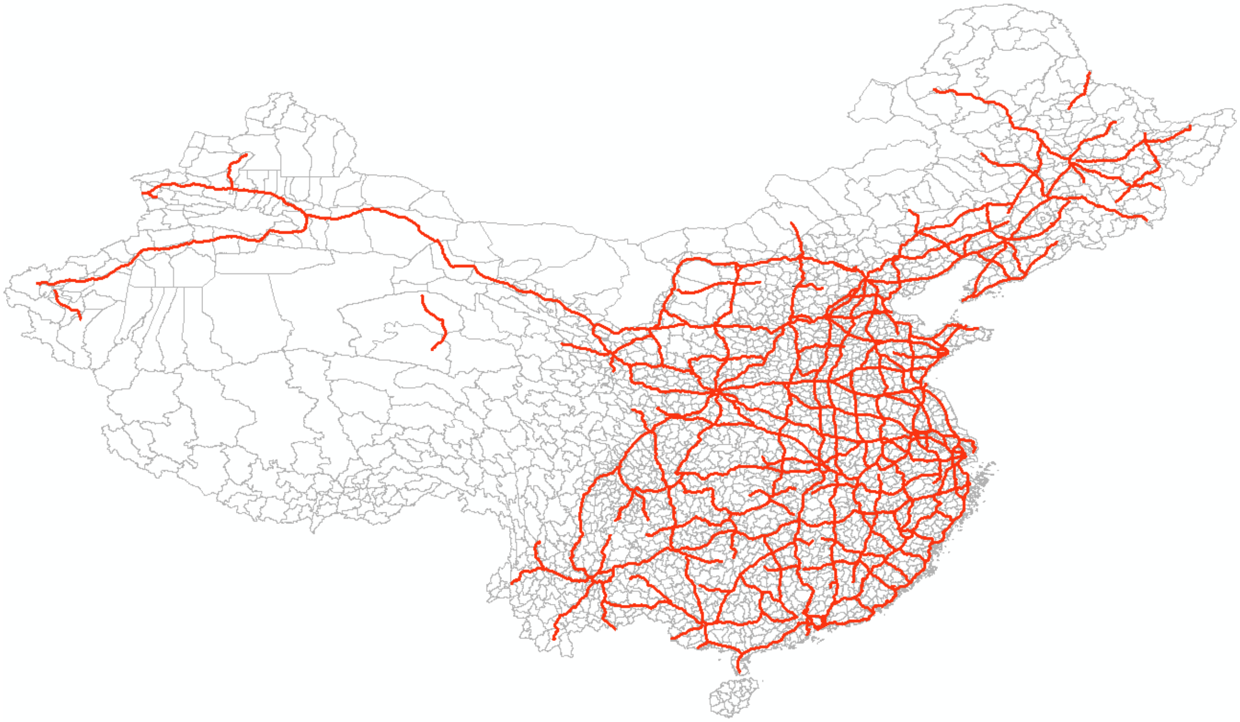
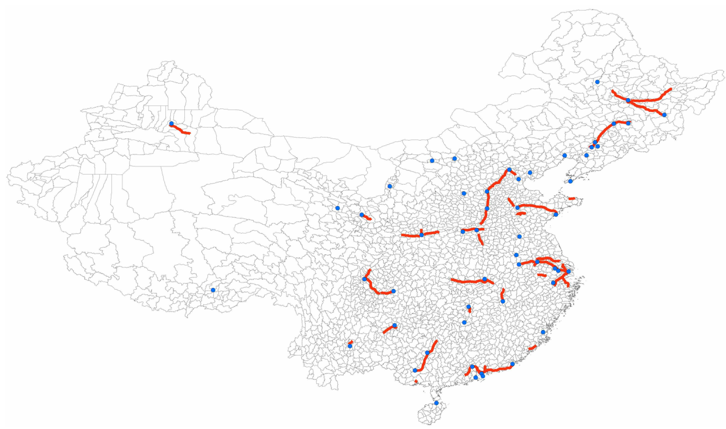
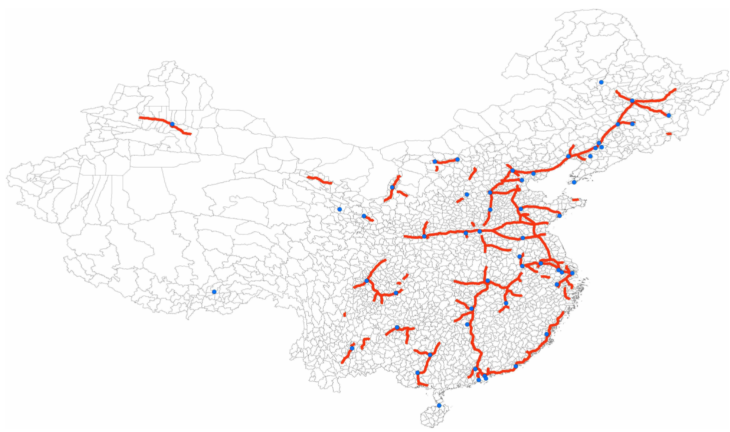


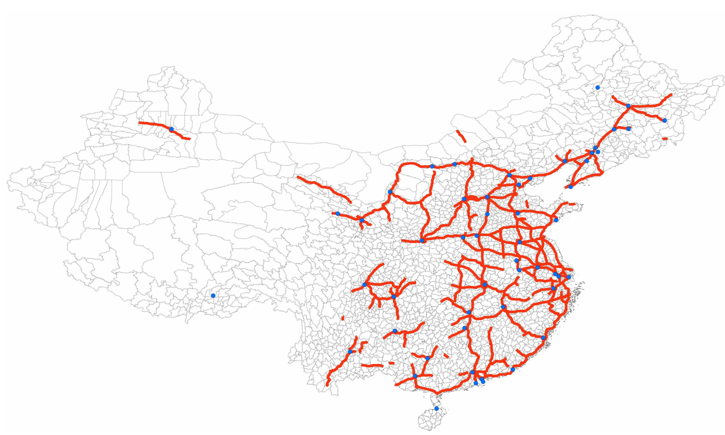
Figure 1
Segments of China's Trunk Highway System Completed by 2013



NTHS in 1999



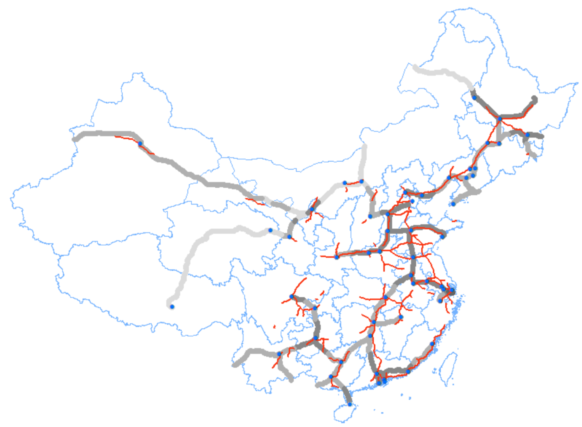
NTHS in 2003



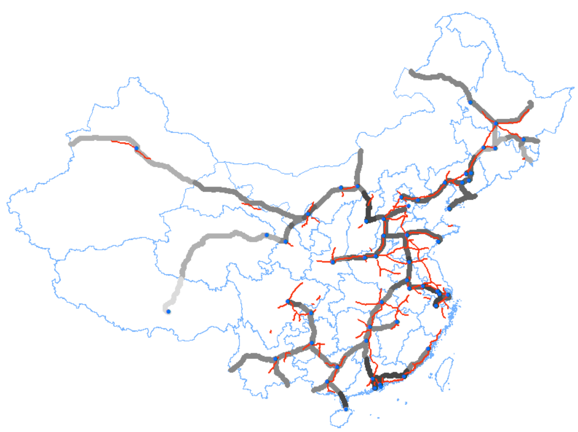
NTHS in 2007

Figure 2
Evolution of China's NTHS

Counterfactual Implementation in 1999



Counterfactual Implementation in 2003



Counterfactual Implementation in 2007

Figure 3

Implementation of Counterfactual Network over Time: The thick grey/black lines represents the counterfactual highway. The shade represents the percentage component, darker means higher percentage. The red lines are the actual highways built.

Table 1
First Stage Regression

Variables	$Connect_{c,t}$	$Connect_{c,t}$
$ConnectIV_{c,t}$	0.562***	
	(0.0459)	
$ConnectIV_{c}$		0.230***
		(0.0162)
Observations	21,690	21,690
R-squared	0.766	0.222
1st Stage F-Stat	150.08	201.38
County FE	Yes	No
Year FE	Yes	Yes
Cluster SE	Yes	Yes

Notes: The unit of observation is county-year. The regressions include county and year fixed effects. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 2
NTHS Connection Effects Among All Firms in the Peripheral Regions

Dependent variables		OLS No Controls	OLS With Controls	IV No Controls	IV With Controls
Δ ln(Output)	Connect	0.0050	0.0037	0.106***	0.0925***
		(0.0048)	(0.0049)	(0.0309)	(0.0308)
	Obs	353,602	353,151	353,602	353,151
	R^2	0.322	0.325	0.321	0.324
Δ ln(Sales)	Connect	0.0048	0.0022	0.106***	0.0868***
		(0.0049)	(0.0049)	(0.0319)	(0.0318)
	Obs	353,540	353,097	353,540	353,097
	R^2	0.314	0.313	0.317	0.315
Δ ln(Profit)	Connect	0.0102	0.0017	0.131	0.107
		(0.0138)	(0.0142)	(0.0905)	(0.0909)
	Obs	332,486	332,472	332,486	332,472
	R^2	0.154	0.155	0.154	0.155
Δ ln(Value-added)	Connect	0.0068	0.0057	0.215***	0.196***
		(0.0074)	(0.0076)	(0.0479)	(0.0480)
	Obs	354,204	353,748	354,204	353,748
	R^2	0.215	0.216	0.213	0.214
Δ ln(ValAdd/Worker)	Connect	0.0086	0.0101	0.262***	0.261***
		(0.0078)	(0.0080)	(0.0508)	(0.0510)
	Obs	353,722	353,423	353,722	353,423
	R^2	0.176	0.173	0.176	0.173

Notes: The unit of observation is firm-year. Each point estimate stems from a separate regression. All regressions include firm and year fixed effects. "With Controls" refers to the regressions that include the log of total assets, a dummy for firms whose majority capital are from State-owned Enterprises and the interaction of the ownership dummy with Connect dummy. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 3
NTHS Connection Effects Among Pre-existing Firms in the Peripheral Regions

Dependent variables		OLS No Control	OLS w/Controls	IV No Control	IV w/Controls
Δ ln(Output)	Connect	-0.0148***	-0.0171***	0.0813**	0.0683**
		(0.0048)	(0.0049)	(0.0317)	(0.0314)
	Obs	299,157	298,717	299,157	298,717
	R^2	0.307	0.310	0.306	0.309
Δ ln(Sales)	Connect	-0.0156***	-0.0190***	0.0832**	0.0637*
		(0.00495)	(0.00507)	(0.0328)	(0.0326)
	Obs	299119	298687	299119	298687
	R^2	0.298	0.301	0.297	0.299
Δ ln(Profit)	Connect	-0.0128	-0.0236	0.0731	0.0438
		(0.0141)	(0.0145)	(0.0938)	(0.0941)
	Obs	280536	280524	280536	280524
	R^2	0.145	0.145	0.145	0.145
Δ ln(ValueAdded)	Connect	-0.0115	-0.0134*	0.201***	0.180***
		(0.0074)	(0.0076)	(0.0491)	(0.0490)
	Obs	299704	299259	299704	299259
	R^2	0.205	0.206	0.202	0.203
Δ ln(ValAdd/Worker)	Connect	-0.0058	-0.0043	0.238***	0.236***
		(0.0078)	(0.0081)	(0.0521)	(0.0522)
	Obs	299,266	298,973	299,266	298,973
	R^2	0.167	0.167	0.163	0.163

Notes: The unit of observation is firm-year. Each point estimate stems from a separate regression. All regressions include firm and year fixed effects. "With Controls" refers to the regressions that include the log of total assets, a dummy for firms whose majority capital are from State-owned Enterprises and the interaction of the ownership dummy with Connect dummy. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4
NTHS Connection Effects Among Pre-existing Firms in the Peripheral Regions on Inputs

Dependent variables		OLS No Controls	OLS With Controls	IV No Controls	IV With Controls
Δ ln(Intermediates)	Connect	-0.0220***	-0.0209***	0.0613**	0.0547*
		(0.0051)	(0.0052)	(0.0299)	(0.0300)
	Obs	298,822	298,822	298,822	298,822
	R^2	0.237	0.237	0.236	0.236
Δ ln(Employment)	Connect	-0.00550	-0.00428	-0.0518**	-0.0549**
		(0.0037)	(0.0038)	(0.0225)	(0.0223)
	Obs	303,050	303,050	303,050	303,050
	R^2	0.221	0.221	0.220	0.220
Δ ln(TotalAssets)	Connect	-0.0285***	-0.0271***	0.0641***	0.0647***
		(0.0052)	(0.0055)	(0.0238)	(0.0245)
	Obs	303,480	303,480	303,480	303,480
	R^2	0.195	0.195	0.194	0.194

Notes: The unit of observation is firm-year. Each point estimate stems from a separate regression. All regressions include firm and year fixed effects. "With Controls" refers to the regressions that include a dummy for firms whose majority capital are from State-owned Enterprises and the interaction of the ownership dummy with Connect dummy. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5
NTHS Connection Effects Among Pre-existing Firms in the Peripheral Regions on Average Wages

Dependent variables		OLS No Controls	OLS With Controls	IV No Controls	IV With Controls
Δ ln(AvgWage)	Connect	0.0004	0.0027	-0.0090	-0.014
		-0.0046	-0.0047	-0.027	-0.0267
	Obs	-0.0046	-0.0047	-0.027	-0.0267
	R^2	0.139	0.139	0.139	0.139

Notes: The unit of observation is firm-year. Each point estimate stems from a separate regression. All regressions include firm and year fixed effects. "With Controls" refers to the regressions that include a dummy for firms whose majority capital are from State-owned Enterprises and the interaction of the ownership dummy with Connect dummy. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 6
NTHS Connection Effects Among Counties in the Peripheral Regions on Imported Inputs, Capital and Intermediates

Dependent variables		(1) (ArcSinh)	(2) (log)	(3) (Mean growth)
Δ g(Imported Inputs)	Connect	2.055	1.862	0.161
		(1.684)	(1.588)	(0.441)
	Obs	6,532	6,532	6,532
	R^2	0.103	0.105	0.168
Δ g(Imported Capital)	Connect	3.252*	3.078*	0.714*
		(1.970)	(1.867)	(0.430)
	Obs	6,532	6,532	6,532
	R^2	0.075	0.076	0.096
Δ g(Imported Intermediates)	Connect	1.334	1.999	0.361
		(1.576)	(1.823)	(0.447)
	Obs	6,532	6,532	6,532
	R^2	0.108	0.095	0.147

Notes: The unit of observation is county-year. Each point estimate stems from a separate regression. All regressions use instrumental variable approach and include county and year fixed effects. All dependent variables are changes in growth rate. Column (1) computes the growth rates as differences in inverse hyperbolic sines between two consecutive years. Column (2) computes the growth rate as differences in logarithms between two consecutive years and setting the zero values to 1. Column (3) uses the mean growth rate between two consecutive years. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 7
NTHS Connection Effects on the Import of Machinery Among Firms in the Peripheral Regions

VARIABLES	(1) OLS	(2) IV
Connect10	0.212**	0.604**
	(0.0968)	(0.283)
Observations	9,829	9,829
R^2	0.214	0.212
Firm FE	YES	YES
Year FE	YES	YES

Notes: The unit of observation is firm-year. Each point estimate stems from a separate regression. All regressions include firm and year fixed effects. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 8

NTHS Connection Effects on Peripheral Counties' Annual Growth Rate

Dependent variables		OLS No Control	OLS With Control	IV No Control	IV With Control
Δ ln(Output)	Connect	-0.0178	-0.0267	0.0833	0.0524
		(0.0144)	(0.0188)	(0.110)	(0.103)
	Obs	11,993	11,990	11,993	11,990
	R^2	0.157	0.210	0.154	0.209
Δ ln(Sales)	Connect	-0.0142	-0.0225	0.109	0.0811
		(0.0135)	(0.0175)	(0.102)	(0.0934)
	Obs	11,996	11,993	11,996	11,993
	R^2	0.145	0.190	0.141	0.188
Δ ln(Profit)	Connect	-0.0317	-0.0438	0.144	0.105
		(0.0224)	(0.0299)	(0.227)	(0.209)
	Obs	11,942	11,942	11,942	11,942
	R^2	0.101	0.119	0.099	0.118
Δ ln(Value-added)	Connect	-0.0222	-0.0316*	0.122	0.0890
		(0.0141)	(0.0186)	(0.130)	(0.129)
	Obs	11,999	11,996	11,999	11,996
	R^2	0.130	0.170	0.127	0.167

Notes: The unit of observation is county-year. Each point estimate stems from a separate regression. All regression include county and year fixed effects. "With Control" refers to the regressions that include log of total industrial assets. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 9

NTHS Connection Effects on Peripheral Firms' Nine-Year Growth Rate

Dependent variables		OLS No Control	OLS With Controls	IV No Control	IV With Controls
Δ ln(Output)	Connect	-0.105***	-0.120***	-0.278	-0.724**
		(0.0275)	(0.0278)	(0.252)	(0.283)
	Obs	9,621	9,613	9,621	9,613
	R^2	0.063	0.133	0.059	0.081
Δ ln(Sales)	Connect	-0.113***	-0.124***	-0.300	-0.751***
		(0.0278)	(0.0282)	(0.255)	(0.287)
	Obs	9,621	9,613	9,621	9,613
	R^2	0.064	0.130	0.059	0.074
Δ ln(Profit)	Connect	-0.0356	-0.0490	0.322	-0.0502
		(0.0554)	(0.0574)	(0.491)	(0.549)
	Obs	9,169	9,169	9,169	9,169
	R^2	0.035	0.062	0.030	0.060
Δ ln(Value-added)	Connect	-0.146***	-0.155***	-0.337	-0.711**
		(0.0328)	(0.0336)	(0.301)	(0.338)
	Obs	9,628	9,620	9,628	9,620
	R^2	0.060	0.107	0.056	0.078

Notes: The unit of observation is firm. Each point estimate stems from a separate regression. All regression include province fixed effect and log distance to the nearest targeted city. "With Controls" refers to the regressions that include log total assets, a dummy for firms whose majority capital are from State-owned Enterprises and the interaction of the ownership dummy with Connect dummy. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 10
NTHS Connection Effects on Peripheral Counties' Nine-Year Growth Rate

Dependent variables		OLS No Control	OLS With Controls	IV No Control	IV With Controls
Δ ln(Output)	Connect	0.0998	-0.0987	-0.200	-0.704*
		(0.0624)	(0.0639)	(0.338)	(0.423)
	Obs	1,217	1,217	1,217	1,217
	R^2	0.161	0.226	0.145	0.167
Δ ln(Sales)	Connect	0.0751	-0.118*	-0.293	-0.793*
		(0.0627)	(0.0643)	(0.341)	(0.429)
	Obs	1,217	1,217	1,217	1,217
	R^2	0.160	0.221	0.136	0.148
Δ ln(Profit)	Connect	-0.0115	-0.311***	-0.822	-1.648**
		(0.0939)	(0.0958)	(0.523)	(0.665)
	Obs	1,211	1,211	1,211	1,211
	R^2	0.156	0.224	0.102	0.095
Δ ln(Value-added)	Connect	0.0264	-0.156**	-0.208	-0.606
		(0.0711)	(0.0736)	(0.382)	(0.475)
	Obs	1,219	1,219	1,219	1,219
	R^2	0.145	0.193	0.138	0.167

Notes: The unit of observation is firm. Each point estimate stems from a separate regression. All regression include province fixed effect and log distance to the nearest targeted city. "With Controls" refers to the regressions that include log total assets, a dummy for firms whose majority capital are from State-owned Enterprises and the interaction of the ownership dummy with Connect dummy. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 11 – Percentage of Counties Connected by Year

1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
13.9%	15.3%	18.8%	23.8%	27.8%	32.3%	36.6%	42.7%	47.4%	50.2%

A Appendix

A.1 Firm-level Data Details

Descriptive statistics for 1999 and 2007 are shown in following Table A1. Table A2 describes firms reporting in the panel, which could be used to proxy for the firm entry and exit.

A.2 Robustness on the Main Results

This subsection addresses identification concerns about the time-varying component of the instrumental variable used in this paper. Since the time-variation is based on the percentage of planned NTHS highways built in each province, potentially unobserved growth characteristics specific to each province could affect the results. Three additional specifications are added to check for robustness for the first stage regression and the main results. The first additional specification uses linear province trend, the second one controls for percent of highway built and the third one adds province-time fixed effects.

Table A3 presents the first stage regression results for my instrumental variable at the county level with the additional specifications. Even with the provincial level controls, the instrument is still a strong and significant predictor of the actual NTHS placement and timing. Table A4, A5 and A6 present the main results in Section 4.1 with specifications that controls for linear province trend, percent of highway built and province-time fixed effects, respectively. These results show that the main findings are robust to alternative specifications that address unobserved province-specific growth patterns.

Table A1 – Summary Statistics

This table describes the main variables in the firm level dataset. Panel A and Panel B reports mean levels for year 1999 and 2007 respectively. "Targeted" are firms located in a postcode within 50km radius from the center of a metropolitan city. "Connected periphery" are firms located in a postcode within 15km distance from an expressway of the NTHS. "Non-connected periphery" are firms located in a postcode more than 50km away from a metropolitan city and more than 15km away from an expressway of the NTHS. Variables with a superscript * indicates 1,000 RMB. Variable with a superscript * indicates 1,000,000 RMB. The top and bottom 1% of each variable are dropped.

Panel A: 1999

	All	Targeted	Connected periphery	Non-connected periphery
Employment	275.2	286.8	294.6	260.7
Wage per Worker*	7.341	8.972	7.062	5.985
Age	15.36	13.50	14.91	16.95
Sales**	29.32	35.80	32.88	23.18
Total asset**	41.83	53.24	41.37	32.46
Total Profit**	1.823	2.380	1.719	1.378
Value added**	8.385	9.909	9.064	6.968
Value added per capital	1.948	2.167	1.970	1.760
Value added per worker*	40.10	45.48	39.36	35.48
Total debt**	26.99	33.04	26.81	22.04
Number of firms	162,033	69,270	16,541	76,222

Panel B: 2007

	All	Targeted	Connected periphery	Non-connected periphery
Employment	177.8	190.1	167.9	174.8
Wage per Worker*	18.27	20.68	16.62	16.72
Age	9.509	9.385	8.560	9.373
Sales**	67.64	78.59	59.29	63.11
Total asset**	55.24	73.21	43.66	46.14
Total Profit**	4.195	4.892	3.440	4.246
Value added**	19.85	22.05	17.28	20.19
Value added per capital	4.102	4.425	3.888	3.915
Value added per worker*	135.17	139.66	124.51	143.65
Total debt**	29.26	36.93	23.97	24.95
Number of firms	336,768	135,888	118,968	81,912

Table A2 – Firm Entry and Exit - Reporting

This table describes firm reporting in the panel. Column (1) displays the number of firms reporting in a particular year. Column (2) reports the percentage of firms reporting in the previous year but not reporting in the current year. Column (3) reports the percentage of firms in the current year that did not report in the previous year. Column (4) reports the percentage of firms in the current year that were born during the year.

	(1)	(2)	(3)	(4)
Year	Number of firms	No reporting	New reporting	New firm
2000	156,819	16.2%	17.0%	1.7%
2001	164,866	23.6%	27.3%	3.1%
2002	175,259	15.3%	20.3%	1.9%
2003	190,122	17.4%	23.9%	3.5%
2004	267,510	28.8%	49.4%	5.2%
2005	261,940	17.9%	16.1%	3.8%
2006	291,815	10.4%	19.6%	3.4%
2007	325,837	10.1%	19.5%	4.0%

Table A3
Robustness Check for First Stage Regression

	(1)	(2)	(3)	(4)
Variables	$Connect_{c,t}$	$Connect_{c,t}$	$Connect_{c,t}$	$Connect_{c,t}$
$ConnectIV_{c,t}$	0.356***	0.322***	0.279***	0.562***
	(0.0486)	(0.0502)	(0.0513)	(0.0459)
Observations	21,690	21,690	21,690	21,690
R-squared	0.778	0.777	0.784	0.766
1st Stage F-Stat	58.47	266.75	29.64	150.08
Controls	Linear province trend	Percentage built	Province*time FE	NO

Notes: The unit of observation is county-year. Column (1) includes linear province trend, Column (2) includes percentage built and Column (3) includes province-time fixed effects. Column (4) is the baseline result from Table 1. The regression include county and year fixed effects. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A4
NTHS Connection Effects Among All Firms in the Peripheral Regions Controlling for Province Linear Trend

Dependent variables		OLS No Controls	OLS With Controls	IV No Controls	IV With Controls
Δ ln(Output)	Connect	0.000524	-0.00124	0.182***	0.173***
		(0.00472)	(0.00485)	(0.0352)	(0.0353)
	Obs	353,602	353,151	353,602	353,151
	R^2	0.323	0.326	0.320	0.322
Δ ln(Sales)	Connect	-0.000728	-0.00403	0.181***	0.165***
		(0.00483)	(0.00496)	(0.0361)	(0.0363)
	Obs	353,540	353,097	353,540	353,097
	R^2	0.315	0.318	0.311	0.313
Δ ln(Profit)	Connect	0.00971	0.00317	0.0566	0.0471
		(0.0128)	(0.0130)	(0.0942)	(0.0939)
	Obs	332,486	332,472	332,486	332,472
	R^2	0.155	0.156	0.155	0.156
Δ ln(Value-added)	Connect	0.00463	0.00335	0.342***	0.335***
		(0.00683)	(0.00702)	(0.0493)	(0.0494)
	Obs	354,204	353,748	354,204	353,748
	R^2	0.215	0.217	0.210	0.210
Δ ln(ValAdd/Worker)	Connect	0.00717	0.00740	0.369***	0.367***
		(0.00706)	(0.00722)	(0.0514)	(0.0513)
	Obs	353,722	353,423	353,722	353,423
	R^2	0.177	0.177	0.170	0.170

Notes: The unit of observation is firm-year. Each point estimate stems from a separate regression. All regressions include a linear trend, firm and year fixed effects. "With Controls" refers to the regressions that include the log of total assets, a dummy for firms whose majority capital are from State-owned Enterprises and the interaction of the ownership dummy with Connect dummy. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A5
NTHS Connection Effects Among All Firms in the Peripheral Regions
Controlling for Percentage Built

Dependent variables		OLS No Controls	OLS With Controls	IV No Controls	IV With Controls
Δ ln(Output)	Connect	0.000336	-0.000178	0.144***	0.135***
		(0.00471)	(0.00483)	(0.0489)	(0.0483)
	Obs	353,602	353,151	353,602	353,151
	R^2	0.322	0.325	0.320	0.323
Δ ln(Sales)	Connect	-0.000911	-0.00264	0.135***	0.121**
		(0.00482)	(0.00494)	(0.0500)	(0.0495)
	Obs	353,540	353,097	353,540	353,097
	R^2	0.314	0.317	0.312	0.314
Δ ln(Profit)	Connect	0.00996	0.00271	0.231*	0.209*
		(0.0127)	(0.0130)	(0.129)	(0.127)
	Obs	332,486	332,472	332,486	332,472
	R^2	0.154	0.155	0.154	0.154
Δ ln(Value-added)	Connect	-0.000605	-0.00119	0.319***	0.297***
		(0.00681)	(0.00699)	(0.0674)	(0.0666)
	Obs	354,204	353,748	354,204	353,748
	R^2	0.215	0.216	0.210	0.211
Δ ln(ValAdd/Worker)	Connect	-0.00228	-0.00136	0.370***	0.363***
		(0.00703)	(0.00718)	(0.0699)	(0.0690)
	Obs	353,722	353,423	353,722	353,423
	R^2	0.176	0.176	0.170	0.170

Notes: The unit of observation is firm-year. Each point estimate stems from a separate regression. All regressions include a control for percentage of highway built, firm and year fixed effects. "With Controls" refers to the regressions that include the log of total assets, a dummy for firms whose majority capital are from State-owned Enterprises and the interaction of the ownership dummy with Connect dummy. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A6
NTHS Connection Effects Among All Firms in the Peripheral Regions Using Province-Time Fixed Effects

Dependent variables		OLS No Controls	OLS With Controls	IV No Controls	IV With Controls
Δ ln(Output)	Connect	0.00302	-0.00203	0.214***	0.207***
		(0.00492)	(0.00507)	(0.0417)	(0.0419)
	Obs	353,602	353,151	353,602	353,151
	R^2	0.334	0.338	0.330	0.333
Δ ln(Sales)	Connect	0.00284	-0.00363	0.201***	0.191***
		(0.00503)	(0.00518)	(0.0428)	(0.0431)
	Obs	353,540	353,097	353,540	353,097
	R^2	0.325	0.329	0.321	0.323
Δ ln(Profit)	Connect	0.0121	0.000894	0.0935	0.0850
		(0.0135)	(0.0137)	(0.112)	(0.111)
	Obs	332,486	332,472	332,486	332,472
	R^2	0.159	0.160	0.159	0.160
Δ ln(Value-added)	Connect	0.00380	-0.000593	0.365***	0.357***
		(0.00714)	(0.00735)	(0.0581)	(0.0582)
	Obs	354,204	353,748	354,204	353,748
	R^2	0.222	0.224	0.216	0.217
Δ ln(ValAdd/Worker)	Connect	0.00147	0.00115	0.356***	0.351***
		(0.00740)	(0.00758)	(0.0602)	(0.0599)
	Obs	353,722	353,423	353,722	353,423
	R^2	0.182	0.183	0.177	0.177

Notes: The unit of observation is firm-year. Each point estimate stems from a separate regression. All regressions include firm, year fixed effects and province-year fixed effects. "With Controls" refers to the regressions that include the log of total assets, a dummy for firms whose majority capital are from State-owned Enterprises and the interaction of the ownership dummy with Connect dummy. ***, ** and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.