Working paper

International Growth Centre

On the Road

Access to Transportation Infrastructure and Economic Growth in China

Abhijit Banerjee
Esther Duflo
Nancy Qian

December 2010

DIRECTED BY
UNIVERSITY OF OXFORD

FUNDED BY
UKaid from the British people

On the Road:
Access to Transportation Infrastructure and Economic Growth in China

Abhijit Banerjee, Esther Duflo and Nancy Qian*

(Incomplete)

Revised Version: December 2010 [First Version: May 2004]

Abstract

This paper estimates the effect of access to transportation networks on regional economic outcomes in China during 1986-2003. It addresses the problem of endogenous placement of networks by exploiting the fact that these networks tend to connect historical cities. Our results show that proximity to transportation networks have a moderate positive causal effect on per capita GDP growth rates across sectors. While we interpret this as the effect of the transportation network *per se*, it could also be the effect of proximity to a communication line between two big cities. These moderately sized benefits appear to reflect small increases in aggregate production rather than displacement of productive firms to be near transportation networks.

1 Introduction

> "A key issue [on whether railroads benefit economic development], however, is whether such railroad influence was primarily exogenous or endogenous, whether railroads first set in motion the forces culminating in the economic development of the decade, or whether arising in response to profitable situations, they played a more passive role." – Albert Fishlow, *American Railroads and the Transformation of the Ante-bellum Economy*, 1965 pp. 203

*This paper is a work-in-progress and the results should be intepreted with this as a caveat. It updates and supercedes "The Railroad to Success: The Effect of Access to Transportation Infrastructure on Economic Growth in China" (Banerjee, Duflo and Qian, 2004), which used the same basic empirical strategy, but substantially less data. We are grateful to Tom Rawski, Thomas Piketty and the participants at the 2004 MacArthur Network for Inequality Conference in Beijing for comments, the China Summer Institute, for very helpful comments. We thank Zhichao Wei, Gongwen Xu and the large team they assembled for invaluable help in data collection; and Giovanni Zambotti for computational assitance with ArcGIS. Please send comments to banerjee@mit.edu, eduflo@mit.edu, or nancy.qian@yale.edu.

Transportation infrastructure is often mentioned as a key to promoting growth and development. The argument relies on the simple logic that one first needs to have access to markets and ideas before one can benefit from them. This belief is supported by the observation during the developmental process of countries that are now rich such the U.S., Japan and Western Europe, the construction of infrastructure such as railroads occurred during times of rapid economic growth. Today, it is undisputable that richer countries have dramatically better transportation infrastructure than poorer ones. This has caused leading historians such as Christopher Savage to make claims such as "the [historic] role of railroads in the U.S. cannot be overstated".[1] Others, such as Fogel (1962, 1964), take a more skeptical view. He argues that one of the most often mentioned historical innovations in transportation infrastructure, railroads, is less effective for economic development in the United States than pre-existing river networks. Moreover, the policies that drove railroad development ultimately misdirected investment. While Fogel's work does not deny the importance of transportation, it begs the question of whether infrastructure development is worthwhile as an object of policy, or whether it is better to rely on the natural forces of the market and/or competition between local jurisdictions to provide the necessary infrastructure when the demand is there.

This paper is about the impact of access to transportation: Do areas that are "quasi-randomly" assigned to get better access to transportation networks end up with different economic outcomes? Our focus is somewhat different from much of the "causal effects" literature discussed in the next section, which focuses primarily on the role of transportation as a reduction in trade costs and the consequences for trade flows. Our main interest is in the more macro question: do areas that benefit from access to the reduction in trade costs and perhaps other costs, get richer as a consequence? It should be easy to see that this is by no means obvious even if there is clear evidence that trade and other flows such as migration went up when the infrastructure became available. For example, the first consequence of the building of motorable roads in the United States was an increase in the ease with which people could travel out from the cities, but eventually it made wholesale suburbanization possible and left many cities without a viable economic model. While it is entirely possible that this process was good for the country as a whole, it is also possible that it hurt (or more plausibly, that it had some potentially avoidable negative consequences), especially given the importance of agglomeration externalities in urban economics Glaeser and Gottlieb (2009). On the other hand, the same agglomeration externalities potentially make cities the key to economic progress and transportation infrastructure is one of the key factors in creating more cities which then turn into "engines" for promoting growth.[2]

Here, we attempt to empirically examine these two closely related questions. First, we

[1] See *Economic History of Transport* Savage (1966).

[2] See for example the World Bank's World Development Report 2009 on *Reshaping Economic Geography* Aoyama and Horner (2009), for a nuanced statement of this view.

ask whether access to better transportation enrich the *average region* that is so affected (because it draws in or generates more new economic activities) or impoverish it (because it becomes easier for human and physical capital to exit)? Second, we ask whether areas that have better access to transportation networks serve as engines of growth when new economic opportunities arise and growth becomes possible?

Three comments need to be made before a more detailed discussion of our paper. First, these are somewhat long-term questions. We are interested not just in the relatively immediate impact on trade and prices that result from greater access but also the subsequent changes in the pattern of localization of economic activity, as people and factories relocate. Second, the emphasis on the average location in the way we formulate the questions is key – if economic activity becomes concentrated in the bigger cities because of better connectivity, those places will clearly benefit even if others lose out. Finally, there is no hope of providing one definitive answer to these questions, since the answer will clearly depend on the starting point – the first road to connect the agricultural hinterland to a port is very different from the fifth such road.

We use county level economic data from China to try to answer these questions. In many ways, China offers an ideal setting for our work. In the late 19th and early 20th century the Chinese government as well as a set of Colonial powers built railroads connecting the historical cities of China to each other and to the newly constructed so-called treaty ports.[3] We identify our average "treated" areas to be those that were close to the straight line connecting the same set of cities–which is an excellent predictor of where the railroads were built–after excluding the area immediately around the termini, where there is obviously an additional terminus effect. We compare these with areas that are further away from those lines and interpret the result of this comparison as the overall effect of both the railroads and any other transportation infrastructure (like highways) that got constructed in the same corridor (probably for the same reason).

This strategy has a number of advantages. First, it provides us an exogenous source of variation in access to transportation networks. Second, this variation goes back to at least fifty years before our study begins in 1986, by which time the patterns of economic activity would have had ample chance to relocate. We can therefore ask what the long run level effect of being close to the line (and hence to transportation) was, say around 1986. Third, our study period 1986-2003 coincides quite well with China's opening up and subsequent growth acceleration. Our treatment areas were in a very good position to take the lead in exploiting these new opportunities. We therefore also study growth effects of being close to the line over the period 1986-2005.

The results show that being close to the line had a positive level effect. Per Capita GDP was higher in places closer to the line. However the effect is not large. The elasticity of

[3] For example, see Pong (1973).

per capita GDP with respect to distance from historical transportation networks is approximately -0.07. The lack of a large level effect is consistent with independent data from a higher-quality household survey, the *National Fixed Point Survey* (NFS) collected by the ministry of agriculture, which finds no significant effect on average household income. The estimated correlation between log distance and log median household income is -0.04 and statistically insignificant. This might reflect the effects of the two contending forces that we discussed earlier. On the other hand, the growth effect of being close to the line is a precisely measured zero. The correlation between log distance and per capital GDP growth is -0.002 with a standard error of 0.003. Places close to the line grew exactly as fast as places further away. Our data shows that per capita GDP grew 9.8% per year. For counties close the transportation network, defined as having distance that is less than the distance in the sample mean, the growth rate was 10% per year; for counties far away, the growth rate was 9.3% per year. The fact that better access to transportation networks does not have a large impact on the (relative) economic performance of those areas, could be evidence for the Fogelian view that transportation infrastructure by itself does not really do very much, unless there is already a demand for it. China scholars have criticized public investment in transportation infrastructure in China after 1990 on similar grounds Huang (2008).

However, in this paper we also explore an alternative complimentary interpretation. Specifically, we argue that the evidence is also consistent with a model where better transportation does lead to substantial cost savings, but its effects are limited by the lack of factor (capital/skills/management) mobility within China. In particular, this model has the prediction that the impact of transportation infrastructure is likely to be limited precisely under the conditions when inequality will be higher in better connected areas and we do find that inequality does have this pattern. We also examine the model's prediction about the allocation of capital using a firm level data set constructed from four waves of the *Census of Manufacturing Firms* (1993, 2004-06). Consistent with the model, we find that distant counties contain fewer firms.

One concern with this interpretation is whether it is too China-specific to be interesting. After all, China has invested massively in infrastructure such that, at the margin, access to railroads could have been less important. It also has a government well known for its attempts to control mobility.[4] Note, however, that in 1986, there were very few cars in China. Therefore, with the exception of river or coastal routes, almost all people and freight presumably traveled by rail. The few highways that existed were mostly through areas that were close to line. Therefore, overall difference in access to transportation in say 1986, was actually quite substantial. The population nearest to the historical transportation networks (the closest decile of counties to the lines that we draw) are a third of the distance to the nearest railroad as the population furthest away (the furthest decile of counties to the lines

[4]For example, see West and Zhao (2000) for a review of studies on labor migration and Huang (2008) for a discussion on infrastructure investment in recent years in China.

that we draw). In terms of access to highways, the nearest counties had fourteen times the length of highways as the furthest counties.

As far as labor mobility is concerned, the main controls were on the movements of low paid unskilled workers. Skilled workers could typically obtain permission to move from employers. Since the biggest cost of moving illegally was the loss of public services such as subsidized housing, schooling and healthcare, the migration restrictions were also not binding for those who had sufficient wealth. Note that there was no formal law against internal capital mobility. Therefore, even if labor could not move to be with capital, capital could move to be with labor Meng (2005); Meng, Junankar, and Kapuscinski (2004); Meng and Kidd (1997); Meng and Zhang (2001, 2010). Our model therefore starts from limited labor mobility and argues that to explain the observed patterns we need limited capital mobility as well. Such limited mobility might reflect the institutional features of the Chinese economy, but is by no means unique to it.[5]

Our paper contributes to the growing number of recent papers on the impact of transportation infrastructure. Michaels (2008) looks at the effect of highway construction in the United States in the 1950s, using both a difference-in-difference approach and based on the observation that highways tended to be built in either a North-South direction or an East-West direction starting from a big city. Donaldson (2010) studies the effects of railroad construction in 19th century India using a difference-in-difference approach and Keller and Shiue (2008) uses a similar approach to look at the opening up of railways between regions of Germany. All these papers start from a trade framework: The effect of transportation infrastructure is studied from the point of view of market integration. The focus is on price convergence and changes in the relative price of factors, along the lines predicted by trade models. Their results suggest that transportation infrastructure favors greater price convergence and that factor prices shift in the direction predicted by trade theory.

Our study differs in that we are more interested in the long-run impact of infrastructure on macroeconomic outcomes such as per capita GDP and per capita GDP growth. Our estimates provide a much more reduced form effect, which presumably includes not just the possible gains from more efficient trade but also the effects of greater factor mobility, better access to education, health care and finance, and other, more diffuse, effects coming from the diffusion of ideas, technologies, etc. Along these lines, our study is more closely related to Atack, Bateman, Haines, and Margo (2009), who focus on the effect of railroads on urbanization and population growth in the United States. While they primarily use a difference-in-difference approach, they also construct an instrument for the distance to the railroad based on the straight line between the start and end points of a railway line.[6] They find a strong effect on urbanization but a small effect on population growth. Similarly, Faber (2009) follows our study and uses the straight-line instrument to study the impact

[5]For example, see Duflo (2004) for evidence of limited capital mobility within Indonesia.
[6]They cite an earlier version of our paper as the source of the instrument.

of highways in China on economic performance. We discuss the difference between our approaches in detail later in Section 6.3.

Moreover, our paper underscores the ways in which the effects of infrastructure are mediated by the quality of factor markets. This is particularly important in developing countries today as there are often institutional failures which limit the internal mobility of capital and goods. For example, the lack of contract enforcement can make it difficult to reallocate the property rights of large machines across wide spaces. The result that access to infrastructure during a period of rapid economic growth does not cause differential growth rates in China, where institutional factors limit the movement of capital, provides a stark example of the importance of factor mobility.

The plan of the rest of the paper is as follows: We start with a brief review of the literature. Section 3 presents the theoretical framework that we use to think about our results including a simple model of industrial location choice. Section 4 provides the background and the empirical strategy. Section 5 describes the data. Section 6 presents the results. Section 7 offers concluding remarks.

2 Growth, Capital and Mobility

This section briefly discusses factor mobility in China in the aspects that relate to the simple model we present in the next section. We aim to make four points. First, central planning policies caused the endowment of human and physical capital to be higher in urban areas relative to rural areas in the pre-reform era (1949-76). However, to promote rural industrialization, the pre-reform government invested high quantities of capital in rural areas Unger (2002). Second, restrictions on migration largely prohibit the mobility of unskilled labor during the post-reform period of our study and limited financial development probably did inhibit capital mobility West and Zhao (2000). Finally, the post-reform era was characterized by very high growth rates.[7]

Chinese central planners have always focused on economic growth and industrialization. In the early 1950s, this meant moving skilled workers and machines into cities. During this period, the percentage of government revenues used to fund industrial development increased from 32% in 1952 to 57% in 1957 Eckstein (1977). Much emphasis was also put into improving human capital in cities. In addition to moving skilled workers into cities, a special emphasis was put on secondary and higher education. Note that all technical and vocational colleges, and universities in China have always only been in cities. This naturally causes human capital to be drawn into cities even if some of the students were born in rural areas.

Rural areas also received investment, albeit less than the cities. An enormous number

[7]See, for example, Hu, Khan, and Fund. (1997) for an overview of Chinese growth.

of primary school were established so that all rural children will have access to a basic education. Literacy rates in China reportedly improved from less than 20% in 1949 to 68% by 1982, even though almost 80% of the population was still rural Jowett (1989). Rural areas also received investments in physical capital: villages were collectivized and physical capital was owned and managed by collectives. When China de-collectivized during the early 1980s, collective assets were inherited by villages, and were often used to form Town and Village Enterprises (TVE). The ownership structure of TVEs are peculiar to China. For our study, it is important to note three facts. First, a significant proportion of industrial output in China came from TVEs. As a percentage of national industrial output, output from TVEs grew from 9% in 1978 to 36% in 1993.[8] Second, TVE assets are jointly owned by all community residents, which were approximately 200 households in an average village and 3,500 households in an average township. Households owned equal shares in TVEs and it was illegal to sell or transfer their shares to non-community members. Third, the law required that at least 60% of the profits be retained in the village.[9] The data shows that over half of the profit was re-invested.[10] These three facts together suggest that a significant amount of productive capital was in rural areas, and policy both prevented their mobility to cities and promoted further capital accumulation in the rural areas.

Labor mobility was probably even more restricted than capital mobility. If a worker moved without official permission, she lost access to all public goods. For urban residents, this meant the loss access to schools, healthcare, and during the 1980s and early 90s, it also meant the loss of food rations and housing. For rural residents, this meant the loss of farmland. Government permission was easier to obtain for skilled workers such as college graduates who could obtain jobs that assisted them in getting the permission or workers with skills that are needed in specific industries. But for the rest of the population permission was extremely difficult to obtain Meng (2005); Meng, Junankar, and Kapuscinski (2004); Meng and Kidd (1997); Meng and Zhang (2001, 2010). Therefore, while the number of migrant workers increased greatly during this period, most of them were temporary migrants who maintained their original residences.[11]

Finally, it is important to point out the differences in growth rates between cities and rural areas and how they changed over time in China during the post-Mao reform era, when income increased rapidly for the country. During the first years of the period, 1978-84, the real income of rural residents grew at 17.7% per year while it was only 7.9% for urban residents. This pattern was reversed in the mid-1980s and the urban advantage increased steadily for the remainder of the reform era. On average, rural real income growth rates

[8]See the *Statistical Material of Township and Enterprises*, 1992.

[9]See Articles 18 and 32 in *The Regulation on Township and Village Collective Enterprises of the People's Republic of China* (1990).

[10]See *Statisitical Survey of China*, 1992: pp. 67.

[11]There has been numerous studies on migration in China. Yaohui (1999) provides a survey of recent evidence.

declined to only 4.1% while urban real income growth was approximately 6.6% Cai (2010).

3 Conceptual Framework

There are a number of reasons why good transportation infrastructure can be advantageous for economic development. First it plausibly reduces trade costs and promotes market integration. This should reduce price volatility and reallocate resources along the lines of comparative advantage. It also increases market size which allows firms to capture gains from increasing returns and promotes more intense competition. Second, it promotes factor mobility. It is easier to migrate to the city if one can return easily whenever needed. It is easier to lend to a borrower whose project you can visit. It is easier to deposit your savings in a bank if the bank is more accessible. Third, it is easier to take advantage of opportunities for investment in the human capital: you can send your child to a better school or take him to a better doctor. Finally, there are intangible benefits. For example, freer movement of people and goods may bring with it new aspirations, new ideas and information about new technologies.

3.1 A simple model of trade and factor mobility

The goal of the model is to look at the effects of distance in a setting where distance affects both the mobility of goods and that of factors of production. The model will illustrate how access to infrastructure can produce very different results depending on which of the two is more affected by distance.

3.1.1 Building blocks

There are $M + N + 1$ regions in this economy: M distant regions, N connected regions and 1 metropolis. Each region produces one good exclusively for export which could be the same as or different from the goods that it import (e.g. food), and another good which it consumes. These goods could be the same or different, but we assume that the relative price of the exportable in terms of the importable in the "world market" is the same – p. However, distance to the market adds to the cost. We model this by assuming that this transportation cost is increasing in distance from the market such that the price received by the exporters is p in the metropolis, $p(1 - d_1)$ in the connected region and $p(1 - d_2)$ in the distant regions, where $d_2 > d_1$.

Production is carried out by a population of firms of identical size in each region. Production requires two inputs which we will call labor and capital, but could also be labor and human capital with small adjustments in the arguments. Output of the exportable is given by $AK^{\alpha}L^{1-\alpha}(\overline{K})^{\beta}$ everywhere, where $\overline{K}$ is the average level of K in firms in that region.[12]

[12] We could easily let A vary across the locations to captures differences in the flow of ideas.

In other words, in the urban economics tradition, we allow for spillovers from co-location. However we assume that the spillovers are not so large as to swamp diminishing returns entirely: $\alpha + \beta < 1$.[13]

The key assumption is with respect to factor mobility. We assume that labor does not move: The city has an endowment of labor of L while all other regions have an endowment of $L' < L$. Capital, on the other hand, does move, but moving is costly. We assume that in equilibrium, the direction of movement that would be needed is from the various regions to the metropolis. This is consistent with the view that in the initial years of Chinese growth after 1978 a lot of the growth and capital accumulation was in rural areas, and it was only later that economic freedoms were extended to urban areas and the urban growth rate crossed its rural equivalent. Therefore when the rental rate for capital in the metropolis is r we assume that the opportunity cost of capital in the connected regions is $r(1-\rho d_1)$ and that in the distant regions is $r(1-\rho d_2)$. In other words, the further you are, the more it costs you to send capital to the metropolis and therefore you are willing to accept a lower return on capital if it is invested in your own region (because say you can monitor the borrower more easily).[14] We assume that there are no other constraints on mobility (no within region credit constraints for example).

3.1.2 Analysis

Analysis of this model is straightforward. Profit maximization with respect to the inputs yields the generic conditions

$$w = p(1-d)A(1-\alpha)(\frac{K}{L})^{\alpha}(\overline{K})^{\beta} \text{ and} \quad (1)$$
$$r(1-\rho d) = p(1-d)A\alpha(\frac{L}{K})^{1-\alpha}(\overline{K})^{\beta}$$

where w is the wage rate in that type of region, L is the labor endowment, K is the equilibrium amount of capital invested in a firm in that region and d is the corresponding distance variable ($d = 0$ for the metropolis, $d = d_1$ for the connected regions and $d = d_2$ for the distant regions). In addition, there is the capital market clearing condition

$$MK_D + NK_C + K_M = K, \quad (2)$$

where K_D is the average amount of capital used in the distant region (per firm), K_C is the same thing in a connected region and K_M is that in the metropolis. K is the total supply of capital in the economy.

[13] See Duranton and Puga (2004) for a review of this literature.

[14] The equivalent assumption for human capital would be that there is a cost to relocating from your region to the city, but the cost is lower if you are better connected (travel back and forth is easier etc.).

Manipulating the capital demand condition and using the fact $K = \overline{K}$ and $L = L'$ yields

$$\overline{K}^{1-\alpha-\beta} = \frac{p(1-d)}{r(1-\rho d)} A\alpha (L')^{1-\alpha} \tag{3}$$

which tells us that whether the distant regions or the connected have more capital per firm depends on whether the ratio $\frac{(1-d)}{(1-\rho d)}$ is increasing or decreasing in d. If $\rho > 1$, which is the case where capital is less mobile than goods, then the distant region will actually have more capital per worker. Using the wage-rental ratio as the measure of inequality, as is conventional in trade models, we see that

$$\frac{w}{r(1-\rho d)} = \frac{(1-\alpha)(\frac{\overline{K}}{L'})}{A\alpha}. \tag{4}$$

It follows directly that inequality is higher wherever $\overline{K}$ is lower. In other words, if capital is less mobile than goods, then the more distant region would have less inequality because it is able to retain more of its capital. A similar result would hold if we replaced capital by human capital and used the skill premium to measure inequality.

Finally comparing outputs per worker/capita,

$$y = p(1-d)A((\frac{1}{L'})^{\alpha}(\overline{K})^{\alpha+\beta} \tag{5}$$

which can be written as

$$y = p(1-d)A((\frac{1}{L'})^{\alpha}(\frac{p(1-d)}{r(1-\rho d)} A\alpha (L')^{1-\alpha})^{\frac{\alpha+\beta}{1-\alpha-\beta}}. \tag{6}$$

In the case where $\rho < 1$, this expression is clearly decreasing in d since both the $p(1-d)$ term and the $\frac{p(1-d)}{r(1-\rho d)}$ term go down with d, but when $\rho > 1$, we might actually observe the reverse, especially when spillovers are large ($1-\alpha-\beta$ close to zero) and therefore $\frac{\alpha+\beta}{1-\alpha-\beta}$ is large.

3.1.3 Results

The analysis in the last part of the previous sub-section suggests that while it is likely that the more connected region will be richer, the gap may be small if capital mobility is low relative to goods mobility, and spillovers are significant. In this case, inequality would be higher in the more connected area. On the other hand, if goods are less mobile, then the more connected area will be richer for two reasons: Higher prices and more capital per worker.

What is the effect of trade opening in this economy? If we model it as an increase in p, it increases incomes everywhere at the same rate. The rate of growth will not depend on the location.

However this is a bit of an artifact of the way the model is set up. Suppose we add an alternative production technology, which uses only labor and produces a perfect substitute for the importable (think of this as agriculture) using the technology $x = BL$, where L is the labor input. The good is consumed in the location and does not need transporting. The point is that the wage in the exporting sector, w, needs to be bigger than B for there to be production of exportables. Since the wage is typically going to be lower in the more distant region (for the same reason that the per capita output is typically lower), if any region does not produce the exportable, it will be the distant region, which would then specialize in "agriculture".

Now suppose p goes up. Then unless the pattern of specialization shifts, there will be growth in the connected areas but not in the more distant regions. On the other hand, the pattern of specialization could shift, in which case both regions will enjoy a growth spurt.

When is the first scenario, where the distant regions do not participate in the "world economy" and therefore do not share in the growth that the connected regions enjoy, more likely? The answer is obviously when output per capita in those regions is low relative to the connected areas, which is when capital is more mobile than goods or if capital is less mobile, when spillovers are weak.

In other words, a pattern where inequality is higher in more connected areas, but output level differences are small and growth rate differences are absent, is consistent with a setting where capital is less mobile than goods.

4 Historical Background and Empirical Strategy

4.1 The Birth of Modern Infrastructure

As explained above, the basic idea behind our empirical strategy it to examine the correlation between the distance to the nearest straight line connecting two historical cities and the outcomes of interest. Throughout the paper, we assert that these lines capture major transportation networks during the 1980s because they capture the first modern infrastructure (e.g. railroads) built in China and much of the infrastructure development afterwards began by initially building along these routes. Later in Section 6.3, we will provide evidence for our assertion.

To draw the lines, we start with the set of important historical cities in China *circa* 1860: Beijing, Taiyuan, Lanzhou, Xi'an, Chengdu, Guiyang, Kunming and Nanchang. These were urban centers that were politically and economically important. To these we add the four treaty ports that were set up by the League of Eight Nations after they defeated the Qing government in the first opium war in 1842 (Shanghai, Ningbo, Fuzhou and Guangzhou).[15]

[15] The Treaty Ports were established in Article 2 of The Treaty of Nanjing, which was signed between

These four cities were chosen for their strategic locations. Shanghai and Ningbo are on the northern and southern mouth of the Yangtze River, Fuzhou was on the southern coast of the Yellow Sea, and Guangzhou was on the Xi River, near its mouth on the South China Sea. All of these ports were easily accessible by the naval gunships of the Western countries and therefore allowed them to both impose their military presence as well as control international trade with China.

It is important to note that with the exception of perhaps Guangzhou, these Treaty Ports were not historically important for the Chinese trade or political governance. China had conducted a very limited amount of international trade since the 16th Century through the Ming and Qing Dynasties. Similarly, it did not have an outgoing navy until the Opium Wars. Therefore, places such as Shanghai, Ningbo and Fuzhou, which were all inhabited previous to 1842, were no more than rural agricultural areas with small stations for domestic naval patrol boats. Their insignificance before 1842 is shown by the fact that none of the four cities were connected to the Grand Canal, which was a north-south canal built to connect Beijing to the important Southern cities. It follows then, that when we draw lines to connect the Treaty Ports and historical Chinese cities, we are unlikely to be systematically capturing important routes from before 1842. Instead, the lines will capture modern transportation networks built afterwards.

The first and perhaps most important transportation infrastructure are railroads. They were mostly built during the early 20th Century jointly by the Qing government and Western countries. The latter provided much of the financing and had much influence over the placement of the railroads. They were largely built to promote Western economic and military interests in China and connected Treaty Ports to historical cities, and also connected historical cities to Colonial cities outside of China. For example, the British planned and financed railways to connect the Yangtze River valley as well as a north-south railway to connect Wuhan to Guangzhou, against the protest of the Qing government who feared the fast troop deployment from Shanghai and Ningbo that would allow the British. The French planned and financed a railway to connect Kunming to Hanoi, an important city in French Indochina. The Russians planned a railway that was almost a straight line from Beijing to Vladivostok through Liaoning, Jilin and Heilongjiang provinces (Spence, 1991: pp. 249-56).

4.2 Straight Lines

We construct our independent variable using a simple algorithm. We draw a straight line from each historically important city to the nearest Treaty Port and/or to the nearest other

the British and the Qing government. Article 2 requested the four cities we mention and Xiamen to be established as Treaty Ports. But in practice, it seemd that Xiamen did not recieve significant investment from the West and only became seen during the second wave of Treaty Port relinquishment by the Qing in 1865. The other Treaty Ports of the second wave were Tianjin, Niuzhang, Yantai, Zhenjiang, Hankou, Shantou, Taibei and Tainan Spence (1991); Pong (1973). Therefore, in our line construction, we omit Xiamen.

historically important city. If there are two cities (or ports) where the difference in distances are less than 100km, we draw a line to both. The line is continued past the city until it hits a natural barrier (e.g. Tibetan Plateau, coast line), or a border to another country. If extended, many of these lines will reach important Colonial cities outside of China. They are shown in Figure 1.

As expected, the lines drawn this way coincide well with railroads constructed during the early 20th century.[16] The three places where they do not match up with the railroads are North Western China (Xinjiang province) and Tibet, where construction occurred under the Communist government after the 1970s as a part of an attempt to politically integrate those areas into China, and North Eastern China (Manchuria), where most of the construction was done by a *de facto* colonial Japanese government during the 1920-30s as part of their attempt to lay claims to these areas. For this reason our main sample will exclude Xinjiang, Tibet, Inner Mongolia and Manchuria (Figure 1 shows that in Manchuria, most counties have a railroad).

Our main source of plausibly exogenous variation for access to infrastructure is the nearest distance from the center of each county to this straight line. The centroid of counties are illustrated in Figure 1. Both the centroids and the nearest distance is computed by ArcGIS using a Asia Conical projection. We use geographic distance rather than travel distance measured as kilometers. This line is also our proxy for transportation infrastructure.

To check the line does indeed proxy for transportation infrastructure, we estimate the correlation between distance to the line and distance to railroads. We estimate the following equation where the left-hand side variable, $\ln dist\_rr_{cpt}$, is the natural logarithm of the shortest distance from the center of the county to the railroad.

$$\ln dist\_rr_{cpt} = \delta \ln dist\_line_{cp} + \rho_p + \gamma_t + \varepsilon_{cpt} \tag{7}$$

Distance to the railroads for county c in province p is a function of: the natural logarithm of the distance to the nearest line connection treaty ports and historical cities illustrated in Figure 1, $dist\_line_{cp}$; province fixed effects, ρ_p; and year fixed effects γ_t. Note that in general, this cannot be interpreted as the first stage of a 2SLS estimate of the effect of railroads, since there likely has been other transportation infrastructure between those cities for a long time (even if there is no difference in the prevalence of paved road today).

Our main estimating equation is the following.

$$Y_{cpt} = \beta \ln dist\_line_{cp} + \Gamma X_{cp} + \rho_p + \gamma_t + \varepsilon_{cpt} \tag{8}$$

The outcome for county c, province p and year t, Y_{cpt}, is a function of: the natural

[16] While the railroads suffered much damage during World War II, after the war, the Guomingtang (KMT) and then the Communist (post -1949) governments undertook extensive repairs and construction focused on upgrading the physical structure. They mostly did not alter the course of the railroads.

logarithm of shortest distance to the line for county c in province p, $dist\_line_{cp}$; a vector of county-specific controls X_{cp}; province fixed effects, ρ_p; and year fixed effects, γ_t. The standard errors are clustered at the county level. If proximity to the line is beneficial, then $\hat{\beta} < 0$.

The basic idea behind our strategy is that there is nothing else that is different about being close to the straight line between two big cities. This obviously relies on the terminal cities not being chosen so that the straight line between them would run through same economically important region. This is the reason why we focus on the ancient cities of China and the Treaty ports: The historical cities are both sufficiently far from each other and clearly so much more important than anything between them in the historical era that it is easy to be comfortable with this assumption with respect to them, and as for the Treaty ports, their locations clearly had to do with available harbors rather than what was between them and the historical cities. The "unequal treaties" that were signed between China and the League of Eight Nations after the Opium Wars allowed the Western countries to house their military in the Treaty Ports but not beyond. Therefore, these ports were chosen to be easily accessible by European ships and also to be strategically advantageous for reaching Chinese cities in case of an uprising or war. The four Treaty Ports in our sample are all along the coast or a major navigable river. While these locations were inhabited in historical times, they were villages, not prominent historical urban centers such as Nanchang or Xian. Therefore, the lines that we draw between these Treaty Ports and the historical Chinese cities have no reason to go through regions of particular importance to the Chinese.

There are two caveats. First, being closer to the line will by construction mean that a county is also closer to the terminal cities. Therefore, our baseline specification will control for distance to the terminal cities. Second, the line from some historically important cities to a Treaty Port might follow along a river. In this case, distance from our line will also capture the distance from the river, which presumably captures all kinds of other things. To address this, our baseline specifications will always control for distance to the nearest navigable river.

In addition to these two controls, the baseline estimation will also control for other potentially influential factors. We will discuss and motivate these later in the paper.

Note that it is not clear that we can expand the set of cities being connected (and therefore use more of the data) without running into potential concerns. One issue is the one of endogeneity that has already been raised. Another equally important issue comes from the very nature of the construction of lines. We compare places that are close to a line with those that are further away. The implicit assumption is that moving further away from one line does not bring us closer to a different line. We ensure this by having relatively few lines and using a sample of counties that are not too distant from any line. The maximum distance of any county in our sample from the nearest line will be 336 km. Figure 1 shows that there are only ten lines. A problem occurs when there are many lines, because it is harder to move away from one line without moving closer to another line. In Section 6.3, we will discuss the relationship between our results and results obtained by Faber (2009) who starts from our basic strategy and extends by adding many more lines in order to capture the more recently constructed highways in China.

5 Data

This paper uses data from multiple sources. All maps are obtained in digital format from the Michigan China Data Center. We computed the distance used here using ArcGIS software, calculations, assuming a Conical Projection.

The first outcome measure we examine is county level GDP. These are from the Provincial Statistical Yearbooks from China from 1986-2005. We collected data from all published yearbooks that reported county-level statistics on GDP.[17] The variables which were consistently reported included GDP and population. These data are interesting because they measure production whereas previous studies have mainly focused on prices. However, there are many problems with this data. First, GDP may have been measured using different techniques across provinces and over time. As these methods are not well-documented and the underlying microdata are not systematically available, we have no way of directly correcting for them. Second, not all counties report GDP and those that are reported are not a random sample of Chinese counties. Third, many counties do not consistently report over time, which means that we have an unbalanced panel where attrition is non-random. As with the GDP measures, there is little documentation on the sampling used in deciding which counties report GDP and we can do little to correct for it. Our final sample for GDP contains 353 counties of sixteen provinces: Beijing, Heibei, Jiangsu, Anhui, Fujian, Jiangxi, Shandong, Henan, Hubei, Hunan, Guangdong, Guangxi, Guizhou, Gansu, Qinghai, and Ningxia. Not all counties are reported each year. Hence, our sample comprises of an unbalanced panel with 3,039 county-year level observations. Figure 2 maps the counties in our sample for which we have GDP data.

To address these difficulties, we use two additional higher quality data sets. While they cannot allow us to directly correct for the county level GDP data, they do allow us to check that the estimated effects in these two alternative data sets are consistent to our theory. The first of these are household incomes from the National Fixed Point Survey (NFS) conducted by the research arm of the Ministry of Agriculture. The NFS is a longitudinal survey of about 320 villages and 24,000 households distributed across all continental Chinese provinces. The NFS began in the mid-1980s. The villages were chosen in the early 1980s to be nationally representative. According to the RCRE, there has been no attrition except in the cases of

[17] Hard copies of these books are stored in the National Library in Beijing. Research assistants scanned and entered the data into computer format.

administrative mergers at the village level and deaths at the household level. Villages and households are surveyed every year. The survey used a stratified sampling approach. For each province, it first randomly selects a number of counties, and then randomly selects a number of villages within each county. Households are then randomly selected from each village. For this study, we use household level data on income. The number of surveyed households per village ranges from approximately 100.[18] Income data includes data from home production, agricultural production and wages. The data is aggregated to the county and year level. It is important to note that the RCRE provided us with income for each decile of the village income distribution and the gini coefficient. They did not provide us with average income across all households. Therefore, in the analysis, we will focus on income of the 10th, 50th, 90th percentiles and the gini. Figure 3 maps the counties for which we have NFS data.

The second additional data we use are firm level data from the Census of Manufacturing Firms, which are available for 1993, 2004-2006.[19] This data samples all manufacturing firms in China that have market capitalization of five million RMBs or more. We will examine two outcomes, profits and the number of firms. The data are aggregated to the county and year level. As we have the Census of all manufacturing firms, our sample contains data from over a thousand counties across the 24 provinces of our sample. The are roughly evenly distributed across space, and therefore for brevity, we do not provide a map for these provinces.

The three sets of outcome data are matched to the distance data at the county level. We exclude the autonomous regions of Tibet, Xinjiang and Inner Mongolia both because these provinces are predominantly non-Han (ethnic minorities) and faced different policies and because the railroads constructed in these regions were the results of very different imperatives. For the latter reason, we also excluded the three Manchurian provinces of Heilongjiang, Liaoning and Jilin. We also excluded the former Treaty Ports that are now province-level municipalities (e.g. Shanghai, Tianjin, Chongqing) and other large cities that are on the segments of railways. This is to avoid the results being driven by the end-points, which are obviously on the line and were chosen because they were important to start with.

Table 1 describes the data. Panel A describes the geography variables for the counties that we have GDP data for. On average, these counties are 67 km from the nearest line, 139 km from the nearest segment city, 37 km from the nearest railraod and 184 km from the nearest navigable river. On average, they contain seven kilometers of highway and 85 km of paved motorways. Panel B shows that on average, per capita GDP 2,839 RMB. These are not deflated over time. In the regression analysis, any changes in prices over time will be

[18]See Martinez-Bravo, i Miquel, Qian, and Yao (2010) for a description of this data.

[19]These data are in principle available for other years. However, we were only able to obtain these four years for which we could geographically identify the location of the firm at the county level. This data has been recently used by Hsieh and Klenow (2009).

controlled for with year fixed effects. Approximately 961 RMB is of this from production in the primary sector. This includes agriculture and mining. Approximately 1,243 RMB is from production in the secondary sector, which includes manufacturing. Approximately 900 RMB comes from the tertiary sector, which includes services. GDP growth during this period is highest in the tertiary sector at 15% per annum and lowest in the primary sector at 6% per annum. On average, annual GDP growth is 10% across sectors. Note that the variance in the distance from the nearest river is extermely high –the relative distance of the furtherst to the nearest is much higher for distance to the river than other transportation routes. To address the fact that the effect of distance may be diminishing, we will control for both the distance to the river and its squared term in the analysis.

Panel C describes the data from manufacturing firms. On average, firms have annual profits of approximately 7,819 RMB. The total amount of profits in a county is on average 441,995 RMB per year. We categorize firms into three groups. Publicly owned firms include firms owned by the state and/or collective. Privately owned firms are firms for which some shares belong to non-public entities. In most cases, the state is the largest shareholder. Most of these firms are former State Owned Enterprises (SOEs) that have sold shares to private entities. These firms can be privately held or publicly traded. We refer to these firms as privatized public firms. The final groups of firms are individually owned. These are firms started by individual entrepreneurs. They are typically much smaller than firms in the first two categories. The data shows that average firm profits for the first two categories are six to ten times as large as that for individually owned firms. On average, each county has approximately 117 manufacturing firms, half of which are individually owned. The descriptive statistics on the large number of small individually owned firms is consistent with the observation that during this period, there was a large increase in private entrepreneurship.

Panel D describes the NFS data on agricultural household income. It shows that on average the gini coefficient for within village inequality is 0.28. Household income for the bottom 10th percentile household is approximately 49% of the median household income and 22% of the household income of the 90th percentile household. Income inequality is increasing over time–household income is growing at a higher rate for the 90th percentile household (9.1%), than the bottom 10th percentile households (7.5%).

Table 2 shows the correlation between the distance to the historical lines we draw and access to transportation infrastructure, income and a county's geographic size. We divide the counties for which we have GDP data into ten equal frequency groups according to their distance to the nearest line. Column (2) shows that the distance to the line is fifty times as far for the furthest decile as the nearest decile. Column (3) shows that the distance to the railroad is 2.5 times as far for the furthest than the nearest. Columns (4)and (6) show that the distance to the river and the length of paved roads are not systematically correlated with distance to the line. In contrast, Column (5) shows that the nearest decile have more than ten times the length of highways as the furthest decile. Column (7) shows that GDP per capita of the furthest decile is approximately 40% of the per capita of the nearest decile. However, Column (8) shows that GDP growth is not systematically correlated with distance from the line. Column (9) shows that the further counties are significantly larger in land area, a factor that will be controlled for in the regression analysis by controlling for county size (or in the case of the NFS data, village size).

6 Results

6.1 Lines, Railroads and Transportation Networks

Table 3 shows the estimates of the correlation between the distance to the nearest transportation infrastructure and distance to the nearest line connecting an original Treaty Port to a historically important city or a historically city to another historically important city based on equation (7). Distance is measured in terms of kilometers. Panel A shows that distance from the historical lines are positively correlated with distance from railroads, negatively correlated with the length of highways within a county, uncorrelated with the length of paved roads within a county and distance to rivers. The correlations shown in Panels B and C will be discussed later in this section.

6.2 The Effect of Distance from the Line on GDP

To illustrate the effects of our baseline controls, we first estimate the effects of distance to the line on the log of GDP per capita. In Table 4, we begin with a specification that only controls for province and year fixed effects (see column (1)). In columns (2)-(7), we gradually introduce the baseline controls. The full baseline specification is shown in column (7) (e.g., equation 8). The estimates show that the coefficient for the log distance to the historical line is very stable across specifications. The baseline estimate is statistically significant at the 10% level. It shows that the elasticity between the distance to the line and per capita GDP is -0.07. For the remaining results, we will show only the baseline specification for the sake of brevity (i.e., for all regressions, we control for the distance to segment cities, the distance to the nearest navigable river and its squared term, the length of highways within a county, the total area of the county, the length of paved roads within a county, province and year fixed effects).

In Table 5, we examine GDP per capita and annual growth in per capita GDP by sector. We estimate the reduced form effect of the distance to the line from equation (7).The estimates for the full sample are shown in Panel A. Columns (1)-(4) show that distance to the line is negatively correlated with GDP levels across sectors. However, they are significant for all sectors and the tertiary sector, but only at the 15% level for the primary and secondary sectors. Columns (5)-(8) show that distance from these lines are uncorrelated with GDP growth. The estimates are small in magnitude and statistically insignificant.

Since the effect of being closer to a line is as large for the primary sector as for other sectors, the estimated effect of railroads is unlikely to only reflect a reporting effect. It certainly possible that large firms set up their headquarters in places nearer the railroads so that production which happens elsewhere is reported from the location of the headquarter but is unlikely to be the sole force behind our estimates. Primary industries are mainly comprised of household level agricultural production. These households are unlikely to report their revenues as part of a larger firm.

Another issue is whether the effect of being close to the line is purely a displacement effect generated by firms relocating to being closer to the line (so that the areas further from the line lose the output that the places closer to the line gain) or whether there is a net increase in output. To investigate this issue, we repeat the estimation on a sample where the 10% nearest counties are excluded, and then again on samples where the 20% are excluded. If the full sample results are caused by productive firms relocating to be very near the railroad, then the estimated effect should decrease in magnitude when we omit those groups (since one would expect firms that chose to relocate to the close to the railroad to relocate as close as possible to it).

Table 5 Panels B and C show that this is not the case. Our estimated effects on GDP and GDP growth are similar in magnitude across subsamples.

6.3 Robustness to Additional Lines

One obvious concern with our strategy is whether the lines we have constructed capture the relevant transportation networks. Earlier in this section, we showed that proximity to our lines are positively correlated with proximity to transportation infrastructure such as roads and railroads. However, it is still possible that we have missed relevant transportation infrastructure such as the relatively newly constructed highways. The first wave of highway construction during the post-Mao reform era occurred along historical transportation routes and often followed along railroads. However, beginning in the late 1990s, the government began to build highways in other parts of the country. It would be tempting to include the effects of these roads into our estimates by adding additional lines that connect the terminus cities of these highway networks such as in a recent study by Faber (2009). However, this mechanically introduces a problem in that areas that are near these new lines will be on average further away from the existing lines. Therefore, if the original transportation network positively affected GDP, the estimates with additional lines with confound the effects of distance from the initial transportation network with proximity to the new network. Not surprisingly, when Faber (2009) more than doubles the number of lines, he finds that distance from the expanded set of lines have on average a positive effect on growth.

In Table 3 Panel B, we use Faber's (2009) data, which he kindly shared with us, to examine whether these lines are correlated with transportation infrastructure. Panel B shows that on average, the distance from the expanded set of lines is positively correlated to distance from railroads, negatively correlated to the length of highways and roads, and uncorrelated with distance from rivers. These lines differ from our historical lines in that distance from them is also negatively correlated with the length of paved roads. Since the expanded set of lines contain the historical lines and the newly added lines, these correlations are consistent with the fact that the new lines capture new road networks built away from the railroads.

In Panel C, we examine the correlations of the historical and expanded set of lines with transportation infrastructure in one regression. The correlation between our historical lines and transportation infrastructure are robust to controlling for the additional lines. Therefore, it is not the case that the correlation between our historical lines are driven by the correlation between the additional lines and transportation infrastructure.

In Table 6, we show that our estimates of the effect of historical lines and log per capita GDP are robust to controlling for the additional lines. Columns (3) and (6) show that distance to the historical lines are negatively correlated with GDP per capita when we control for the distance to the expanded set of lines with and without our baseline controls. In contrast, they show that the coefficient for the distance to the expanded set is very small in magnitude and statistically insignificant in all specifications. They also change signs when we introduce controls. These estimates suggests that the historical lines are indeed the relevant lines to study in our context.

6.4 The Effect on Household Income and Firms Placement

Table 5 shows the estimated effects of distance on average household income for agricultural households at the village level. Panel A column (1) shows that distance to the line is negatively correlated with median household income. But the estimate is small in magnitude and statistically insignificant. Column (2) shows that distance is negatively correlated with inequality, measured as the gini coefficient. The estimate is statistically significant at the 1% level. In columns (3)-(4), the estimated coefficients suggest that distance is negatively correlated with the incomes of the poorest household more than with the richest households, but the estimates are statistically insignificant.

In Column (5), we estimate the effect of distance on the annual growth in median household. It shows that distance is uncorrelated with median household income growth. The estimate is small in magnitude and insignificant. In column (6), the estimates show that distant villages experience a slower rise in inequality within villages. The estimate is statistically significant at the 1% level. Columns (7) and (8) provide suggestive evidence that income is growing at a slower rate for richer households when the village is further away from the lines. However, these estimates are not significant.

Table 8 shows that the distance to transportation network is negatively correlated with the placement of firms. The further away a county is, the fewer firms it contains. Columns (2) and (4) shows that this is driven by the placement of publicly and individually owned firms. The fact that there is no effect of the number of privatized public firms (column 3) is interesting to note because this means that the privatization of SOEs did not vary according to distance to the line. This is reassuring in that it tells us that the distance to the line (conditional on our controls) is not capturing systematic differences in political changes or other factors that may be correlated with privatization such as corruption. The estimates are similarly negative when we examine the correlation between distance from the historical lines and the growth in the number of firms. However, they are statistically insignificant. We do not report them for the sake of brevity.

Columns (5)-(9) show the estimates on firm profit. The estimates show that distance is negatively correlated with the profits of publicly owned and individually owned firms. However, they are not statistically significant. Note that for presentation purposes, we scaled log profit by 10,000. Therefore, the correlations are also very small in magnitude. We also examined the effect of distance on returns to capital as measured by profits divided by the value of total capital. These estimates were negative, small in magnitude and statistically insignificant. We do not report them in the paper because of concern over the quality of capital data. Specifically, it is unclear how capital is valued by these firms. Much of the capital is inherited from the state or collectives and one would only know the market value if she observed the market transaction of another similar piece of capital. If further away regions have fewer market transactions such that firms there are more likely to under-value the capital, then our estimate of the returns to capital will be systematically over-stated as we move further away. This is a generic problem in the Chinese data on firm assets.

7 Conclusion

In this paper, we investigate the effects of having access to transportation infrastructure during the two decades after China opened up to trade and market reforms, when it experienced rapid GDP growth. We find that regions closer to historical transportation networks have higher levels of GDP per capita, higher income inequality and a higher number of firms. However, these level differences are small in magnitude and we find no evidence that distance affected income growth during the two decades of rapid economic growth after China opened up its economy to trade and market reforms.

Our results do not contradict the more straightforward Fogelian interpretation or Yasheng Huang's (2008) view that during this period of fast growth, China should not have focused so much on building highways. In particular, our interpretation would suggest that improving factor mobility should have taken priority. The theoretical framework and results from this paper highlight an important point: Factor mobility can play an important role in determining the effects of transportation infrastructure. In the China case, we find that capital is immobile relative to goods. Therefore, although the cities were growing faster than distant areas, this relative immobility prevented the larger spatial differences in growth rates that would have occurred had capital been able to flow easily to the cities.

These results should not discourage those who believe that investment in transportation infrastructure can promote growth. Rather they highlight the importance of other factors in thinking of the effects of infrastructure on economic growth. Moreover, as we noted in the introduction of this paper, without knowing the returns of such investment, one cannot say whether investments in transportation infrastructure ought to be made. Finding credible ways to estimate or even bound the social returns remains a very important next step in this research agenda.

References

Aoyama, Y., and R. Horner (2009): "World Development Report 2009: Reshaping Economic Geography, by the World Bank," *Journal of Regional Science*, 50(5), 1467–9787.

Atack, J., F. Bateman, M. Haines, and R. A. Margo (2009): "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-60," NBER Working Papers 14640, National Bureau of Economic Research, Inc.

Cai, F. (2010): "Chinese rural reform in 30 years: An analysis in terms of institutional economics," *Frontiers of Economics in China*, 5, 150–168, 10.1007/s11459-010-0008-1.

Donaldson, D. (2010): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," Mit working papers, MIT.

Duflo, E. (2004): "The medium run effects of educational expansion: evidence from a large school construction program in Indonesia," *Journal of Development Economics*, 74(1), 163–197.

Duranton, G., and D. Puga (2004): "Micro-foundations of urban agglomeration economies," in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson, and J. F. Thisse, vol. 4 of *Handbook of Regional and Urban Economics*, chap. 48, pp. 2063–2117. Elsevier.

Eckstein, A. (1977): *China's economic revolution*. Cambridge University Press.

Faber, B. (2009): "Integration and the Periphery: The Unintended Effects of New Highways in a Developing Country," Lse working papers, LSE.

Fogel, R. (1962): "A Quantitative Approach to the Study of Railroads in American Economic Growth: A Report of Some Preliminary Findings," *Journal of Economic History*, 22(2), 163–197.

——— (1964): *Railroads and American Economic Growth: Essays in Econometric History*. Johns Hopkins Press, Baltimore.

Glaeser, E., and J. Gottlieb (2009): "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States," *Journal of Economic Literature*, 47(4), 983–1028.

Hsieh, C.-T., and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124(4), 1403–1448.

Hu, Z., M. S. Khan, and I. M. Fund. (1997): *Why is China growing so fast? / Zuliu Hu and Mohsin S. Khan*. International Monetary Fund, Washington, D.C. :.

Huang, Y. (2008): *Capitalism with Chinese Characteristics: Entrepreneurship and the State*. Cambridge University Press, 1 edn.

Jowett, A. J. (1989): "Patterns of literacy in the People's Republic of China," *GeoJournal*, 18, 417–427, 10.1007/BF00772696.

Keller, W., and C. H. Shiue (2008): "Institutions, Technology, and Trade," NBER Working Papers 13913, National Bureau of Economic Research, Inc.

Martinez-Bravo, M., G. P. i Miquel, N. Qian, and Y. Yao (2010): "Accountability under Authoritarianism: The Effect of Electoral Reforms in Rural China," Yale university working papers, Yale University.

Meng, X. (2005): "Introduction: Poverty and labor markets in China," *Journal of Comparative Economics*, 33(4), 641–643.

Meng, X., P. N. R. Junankar, and C. A. Kapuscinski (2004): "Job Mobility along the Technological Ladder: A Case Study of Australia," IZA Discussion Papers 1169, Institute for the Study of Labor (IZA).

Meng, X., and M. P. Kidd (1997): "Labor Market Reform and the Changing Structure of Wage Determination in China's State Sector during the 1980s," *Journal of Comparative Economics*, 25(3), 403–421.

Meng, X., and D. Zhang (2010): "Labour Market Impact of Large Scale Internal Migration on Chinese Urban 'Native' Workers," IZA Discussion Papers 5288, Institute for the Study of Labor (IZA).

Meng, X., and J. Zhang (2001): "The Two-Tier Labor Market in Urban China: Occupational Segregation and Wage Differentials between Urban Residents and Rural Migrants in Shanghai," *Journal of Comparative Economics*, 29(3), 485–504.

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.

Pong, D. (1973): "Confucian Patriotism and the Destruction of the Woosung Railway, 1877," *Modern Asian Studies*, 7(04), 647–676.

Savage, C. I. (1966): *An economic history of transport*. Routledge.

Spence, J. (1991): *The search for modern China*. Norton.

Unger, J. (2002): *The transformation of rural China*, Asia and the Pacific. M.E. Sharpe.

West, L. A., and Y. Zhao (2000): *Rural Labor Flows in China*. Berkeley University Press, Berkeley, California.

Yaohui, Z. (1999): "Leaving the Countryside: Rural to Urban Migration Decisions in China," *The American Economic Review Papers and Proceedings*, 89, 281–286.

Table 1: Means

Variable	Obs	Mean	Std. Err.
A. Geography			
Distance to Line (Historical)	3393	67.37	(1.05)
Distance to Segment City	3393	139.39	(1.40)
Distance to Railroad	3393	37.35	(0.60)
Distance to River	3368	184.53	(2.34)
Length of Highways	3368	7.41	(0.25)
Length of Paved Roads	3368	85.52	(0.82)
Area of County (Sq Km)	3393	1950000.00	1130000.00
B. County GDP			
GDP PC	3393	2839.83	(56.83)
GDP PC Primary	2824	961.67	(17.13)
GDP PC Secondary	2824	1243.63	(34.45)
GDP PC Teriary	2732	900.68	(21.88)
GDP PC Growth	3080	0.10	(0.00)
GDP PC Growth Primary	2466	0.06	(0.00)
GDP PC Growth Secondary	2466	0.13	(0.01)
GDP PC Growth Tertiary	2312	0.15	(0.00)
C. Manufacturing Firms			
Average Profits	4504	7819.66	(2933.36)
Total Profits (within County)	4504	441995.30	(26814.02)
Average Profits of Publically Owned Firms	4433	8267.56	(2989.87)
Average Profits of Privatized Public Firms	2903	13221.20	(1432.23)
Average Profits of Individually Owned Firms (True Private Firms)	2694	1436.49	(72.02)
Total Number of Firms (within County)	4504	117.27	(2.92)
Number of Publically Owned Firms	4433	49.74	(1.40)
Number of Privatized Public Firms	2903	21.07	(0.64)
Number of Individually Owned Firms (True Private Firms)	2696	58.67	(2.39)
D. Agricultural Household Income			
Gini	2261	0.28	(0.00)
HH Income 10th Percentile	2261	4565.98	(67.30)
HH Income 50th Percentile	2261	9329.67	(153.29)
HH Income 90th Percentile	2261	20938.15	(570.63)
Gini Growth	2131	0.002	(0.00)
HH Income Growth 10th Percentile	2131	0.075	(0.01)
HH Income Growth 50th Percentile	2131	0.083	(0.00)
HH Income Growth 90th Percentile	2131	0.091	(0.00)

Table 2: Distance Correlates

Distance to Line Decile	Dist to Line	Dist to RR	Dist to River	Highway Length	Paved Road Length	GDP PC (RMB)	GDP PC Growth	Area (Sq Km)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	4.98	19.03	144.43	14.90	73.42	3917.15	0.11	1.53E+06
1	14.50	31.90	159.14	9.78	91.97	3175.31	0.10	1.84E+06
2	23.41	26.41	237.66	10.11	90.06	3949.51	0.11	1.80E+06
3	33.10	27.08	185.06	8.62	70.43	2316.75	0.10	1.59E+06
4	45.53	38.70	183.24	4.65	82.23	2856.87	0.10	2.09E+06
5	59.87	36.61	185.55	5.82	77.64	3246.01	0.09	1.66E+06
6	72.84	42.75	212.69	7.44	98.36	2985.89	0.09	1.99E+06
7	87.87	42.83	165.20	3.85	83.41	2542.15	0.11	1.92E+06
8	118.29	52.76	210.59	7.07	97.33	1779.39	0.09	2.48E+06
9	208.87	54.78	159.99	1.91	89.63	1668.47	0.08	2.57E+06
All	67.37	37.35	184.53	7.41	85.52	2839.83	0.10	1.95E+06

Table 3: Line Correlates

	Dependent Variables: Transportation Infrastructure			
	(1)	(2)	(3)	(4)
	Ln RR Distance	Ln Highway Length	Ln Road Length	Ln River Distance
	A. Lines of Historical Transportation Networks			
Ln Dist Line (Historical)	0.231	-0.212	0.0279	0.0335
	(0.0613)	(0.0856)	(0.0461)	(0.0523)
Observations	3393	3368	3368	3368
R-squared	0.255	0.153	0.353	0.496
	B. Lines Expanded to Include New Highway Networks (Farber, 2009)			
Ln Dist Lines (Expanded)	0.147	-0.393	-0.0807	0.00892
	(0.0561)	(0.0823)	(0.0267)	(0.0365)
Observations	3216	3211	3211	3211
R-squared	0.249	0.234	0.366	0.487
	C. Historical Lines controlling for Expanded Lines			
Ln Dist Line (Historical)	0.223	-0.128	0.0428	0.0357
	(0.0633)	(0.0794)	(0.0481)	(0.0544)
Ln Dist Lines (Expanded)	0.117	-0.375	-0.0865	0.00405
	(0.0554)	(0.0827)	(0.0268)	(0.0356)
Observations	3216	3211	3211	3211
R-squared	0.276	0.240	0.369	0.488

All regressions control for the area of the county, year and province fixed effects. Standard errors are clustered at the county level.

Table 4: The Effect of Distance on Ln GDP PC

	Dependent Variable: Ln GDP PC						
	(1)	(2)	(3)	(4)	(5)	(6)	(7) Baseline
Ln Line Dist	-0.0885	-0.0665	-0.0652	-0.0666	-0.0575	-0.0672	-0.0716
	(0.0411)	(0.0411)	(0.0407)	(0.0406)	(0.0385)	(0.0385)	(0.0382)
Ln Seg Cities Dist		-0.141	-0.148	-0.161	-0.162	-0.149	-0.127
		(0.0587)	(0.0579)	(0.0618)	(0.0604)	(0.0588)	(0.0594)
Ln River Dist			-0.0106	-0.00976	0.00406	0.609	0.680
			(0.0555)	(0.0556)	(0.0558)	(0.180)	(0.191)
Ln River Dist Squared						-0.0746	-0.0823
						(0.0241)	(0.0254)
Ln Highway Length				-0.0155	-0.000749	-0.00334	-0.0107
				(0.0262)	(0.0269)	(0.0264)	(0.0266)
Ln Road Length					-0.112	-0.110	-0.0420
					(0.0813)	(0.0808)	(0.101)
Ln County Area							-0.147
							(0.102)
Observations	3393	3393	3368	3368	3368	3368	3368
R-squared	0.443	0.453	0.472	0.472	0.481	0.490	0.496

All regressions control for province and year fixed effects. Standard errors are clustered at the county level.

Table 5: The Effect of Distance on GDP PC by Sector

	Dependent Variables							
	Ln GDP PC Levels				GDP Growth			
	All	Primary	Secondary	Tertiary	All	Primary	Secondary	Tertiary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Full Sample							
Ln Line Dist	-0.0716	-0.0555	-0.0794	-0.0739	-0.00162	0.000318	-0.000277	0.00399
	(0.0382)	(0.0389)	(0.0517)	(0.0416)	(0.00293)	(0.00383)	(0.00495)	(0.00314)
Observations	3368	2799	2799	2710	2142	1528	1528	1337
R-squared	0.496	0.379	0.460	0.555	0.063	0.039	0.092	0.084
	B. Omit Nearest 10%							
Ln Line Dist	-0.0857	-0.0887	-0.0821	-0.0787	-0.000466	0.00335	-0.00108	0.00682
	(0.0504)	(0.0568)	(0.0714)	(0.0542)	(0.00372)	(0.00424)	(0.00819)	(0.00569)
Observations	3037	2502	2502	2421	1935	1356	1356	1183
R-squared	0.492	0.376	0.450	0.553	0.066	0.039	0.090	0.100
	C. Omit Nearest 20%							
Ln Line Dist	-0.0974	-0.0897	-0.146	-0.0799	-0.00338	0.00383	-0.00325	0.0111
	(0.0605)	(0.0681)	(0.0875)	(0.0636)	(0.00448)	(0.00531)	(0.0102)	(0.00784)
Observations	2696	2199	2199	2123	1710	1171	1171	1010
R-squared	0.496	0.389	0.449	0.558	0.068	0.047	0.089	0.121

All regressions control for ln distance to the segment city, ln distance to the nearest river, ln distance to the nearest river squared, ln length of highways, ln length of paved roads, ln county area, province and year fixed effects. The growth regressions in columns (5)-(8) also control for per capita GDP levels (by sector) lagged one and two years. Standard errors are clustered at the county level.

Table 6: Historical Lines v. Expanded "New" Lines

	Dependent Variables: Ln GDP PC					
	(1)	(2)	(3)	(4)	(5)	(6)
Ln Line (Historical) Dist	-0.0885		-0.0932	-0.0659		-0.0779
	(0.0411)		(0.0424)	(0.0385)		(0.0404)
Ln Line (Expanded) Dist		-0.0395	-0.0265		0.00576	0.00811
		(0.0284)	(0.0269)		(0.0354)	(0.0345)
Controls						
Ln Seg Cities Dist	N	N	N	Y	Y	Y
Ln River Dist, Ln River Dist Squared	N	N	N	Y	Y	Y
Ln Highway Length	N	N	N	Y	Y	Y
Ln Road Length	N	N	N	Y	Y	Y
Ln Area	N	N	N	Y	Y	Y
Observations	3393	3216	3216	3368	3211	3211
R-squared	0.443	0.468	0.477	0.487	0.499	0.506

All regressions control for province and year fixed effects. Standard errors are clustered at the county level.

Table 7: The Effect of Distance on Rural Household Income

	Dependent Variables							
	Levels				Growth			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	50th Ln HH Inc	Gini	10th Ln HH Inc	90th Ln HH Inc	50th Ln HH Inc	Gini	10th Ln HH Inc	90th Ln HH Inc
	A. Full Sample							
Ln Line Dist	-0.00562	-0.00538	-0.03843	-0.00732	0.00014	-0.00131	0.00047	-0.00200
	(0.02010)	(0.00279)	(0.02422)	(0.02197)	(0.00215)	(0.00046)	(0.00278)	(0.00271)
Observations	1965	1965	1965	1965	1722	1722	1722	1722
R-squared	0.806	0.238	0.500	0.760	0.485	0.034	0.060	0.364
	B. Omit Nearest 10%							
Ln Line Dist	0.00719	-0.00265	-0.03479	0.00919	-0.00083	-0.00097	-0.00288	-0.00312
	(0.03365)	(0.00512)	(0.04110)	(0.04275)	(0.00360)	(0.00094)	(0.00459)	(0.00527)
Observations	1774	1774	1774	1774	1555	1555	1555	1555
R-squared	0.805	0.233	0.487	0.755	0.502	0.039	0.060	0.361
	C. Omit Nearest 20%							
Ln Line Dist	0.02969	0.00384	-0.02985	0.04325	-0.00033	-0.00030	-0.00281	0.00278
	(0.04437)	(0.00683)	(0.05626)	(0.05790)	(0.00498)	(0.00104)	(0.00586)	(0.00670)
Observations	1567	1567	1567	1567	1374	1374	1374	1374
R-squared	0.811	0.252	0.480	0.762	0.490	0.038	0.059	0.350

All regressions control for ln distance to the segment city, ln distance to the nearest river, ln distance to the nearest river squared, ln length of highways, ln length of paved roads, ln county area, province and year fixed effects. The growth regressions in columns (5)-(8) also control for lagged level measures of the depdendent variables one and two years. Standard errors are clustered at the county level.

Table 8: The Effect of Distance on Firms

	Dependent Variables								
	Ln Number of Firms				Ln Firm Proft*				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All	Public	Privatized Public	Individual	Sum All	Avg All	Avg Public	Privatized Public	Avg Individual
	A. All								
Ln Line Dist	-0.05858	-0.04498	0.00801	-0.07318	-7.46310	0.86841	-0.03841	5.38287	-0.21680
	(0.01851)	(0.02101)	(0.02951)	(0.03023)	(32.93742)	(0.84613)	(0.49066)	(4.96234)	(0.23889)
Observations	3239	3173	1883	1775	3239	3239	3173	1883	1773
R-squared	0.634	0.697	0.491	0.688	0.292	0.048	0.027	0.083	0.128
	B. Omit Nearest 10%								
Ln Line Dist	-0.06782	-0.05430	0.01795	-0.07195	-46.46761	1.01524	-0.43719	9.67804	-0.72732
	(0.02756)	(0.02770)	(0.04849)	(0.04690)	(38.58965)	(1.15822)	(1.07706)	(8.07610)	(0.52434)
Observations	2926	2865	1685	1594	2926	2926	2865	1685	1592
R-squared	0.649	0.713	0.478	0.700	0.306	0.051	0.028	0.097	0.151
	C. Omit Nearest 20%								
Ln Line Dist	-0.08263	-0.05519	0.05444	-0.11852	-18.00365	1.49797	0.62456	12.58678	-0.55885
	(0.03858)	(0.03454)	(0.06861)	(0.06609)	(39.24528)	(1.21463)	(1.05865)	(9.14912)	(0.56586)
Observations	2598	2540	1483	1401	2598	2598	2540	1483	1399
R-squared	0.658	0.718	0.470	0.694	0.324	0.054	0.027	0.109	0.141

All regressions control for ln distance to the segment city, ln distance to the nearest river, ln distance to the nearest river squared, ln length of highways, ln length of paved roads, ln county area, province and year fixed effects. Standard errors are clustered at the county level. *Note: For interpretational ease , we scaled ln firm profit by 10,000.

Figure 1: Historical Lines

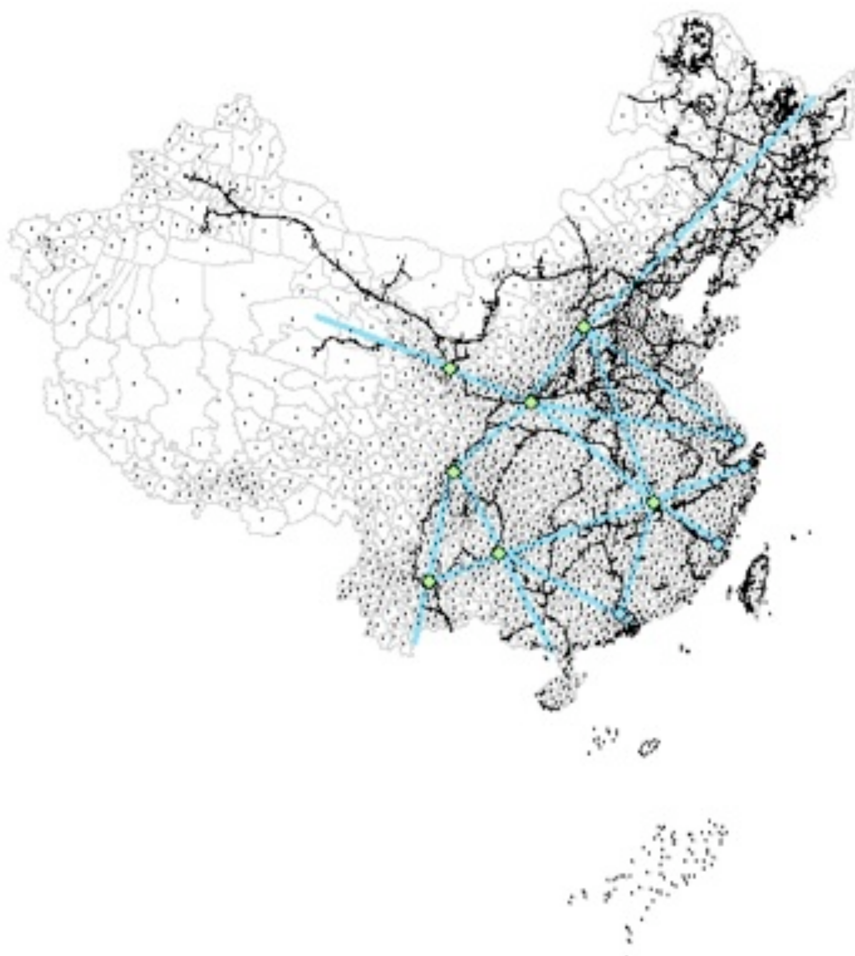


Figure 2: County GDP Data

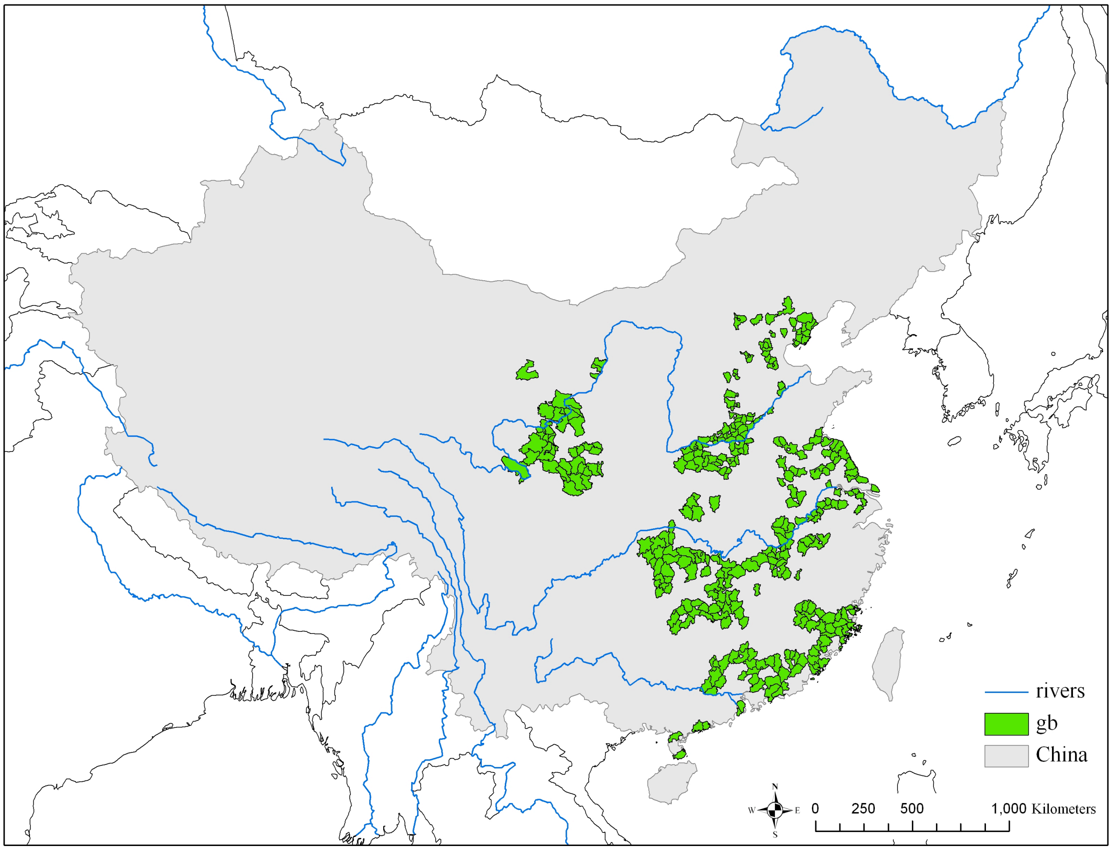


Figure 3: NFS Village Household Income Data

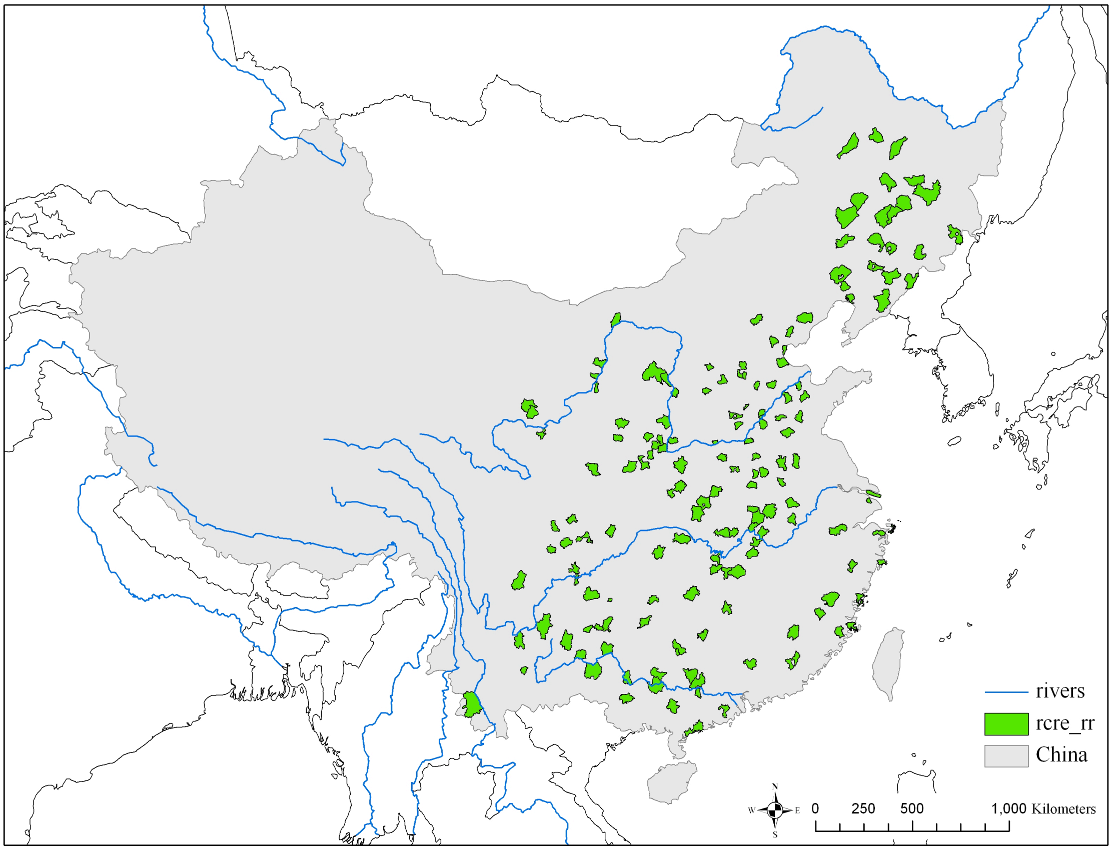
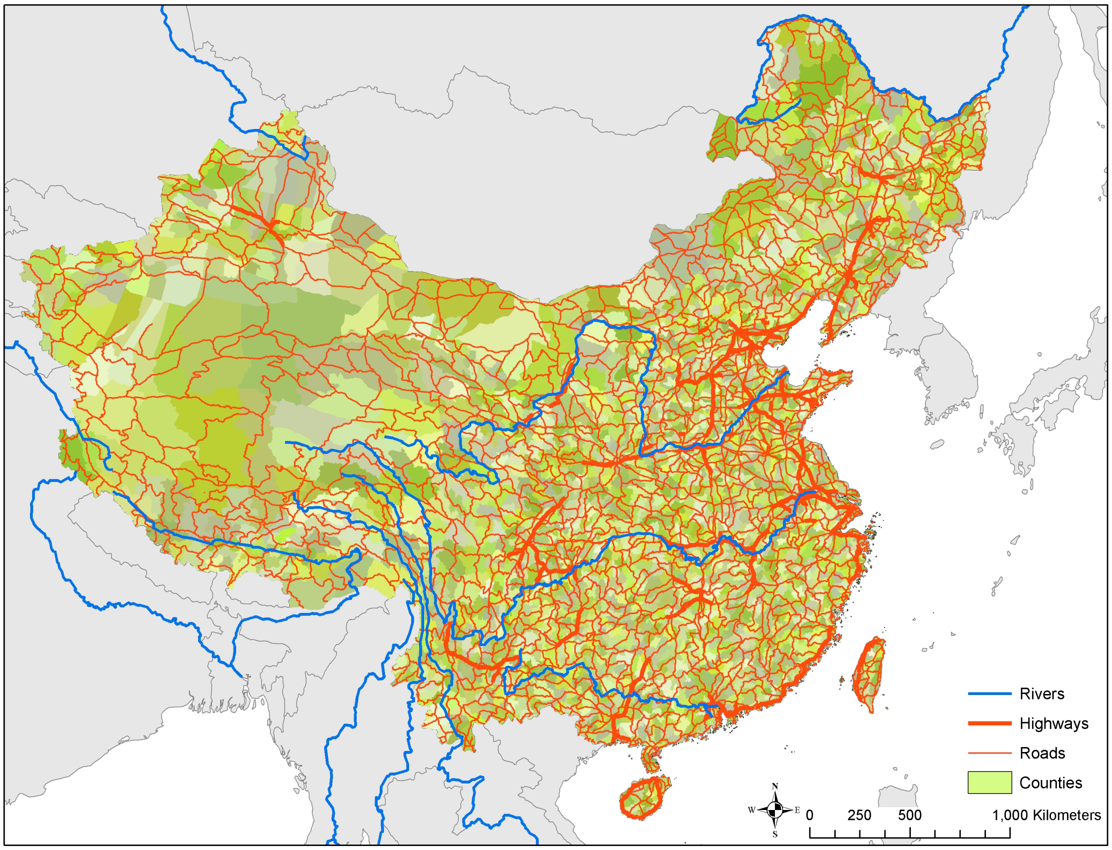


Figure 4: Highways and Rivers



The International Growth Centre (IGC) aims to promote sustainable growth in developing countries by providing demand-led policy advice based on frontier research.

Find out more about
our work on our website
www.theigc.org

For media or communications
enquiries, please contact
mail@theigc.org

Subscribe to our newsletter
and topic updates
www.theigc.org/newsletter

Follow us on Twitter
@the_igc

Contact us
International Growth Centre,
London School of Economic
and Political Science,
Houghton Street,
London WC2A 2AE

International
Growth Centre

DIRECTED BY

FUNDED BY

Designed by soapbox.co.uk