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

Abhijit Banerjee, Esther Duflo and Nancy Qian*

(Preliminary Version)

March, 2009

Abstract

This paper estimates the effect of access to transportation networks on regional demographic and 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 large 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 benefits appear to reflect increases in aggregate production rather than displacement of productive firms to be near transportation networks.

*This paper updates and supercedes "The Railroad to Success: The Effect of Access to Transportation Infrastructure on Economic Growth in China" (Banerjee, Duflo and Qian, 2006), 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. 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

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 Antebellum Economy*, 1965 pp. 203

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 before one can benefit from them. This belief is supported by the observation during the developmental process of countries that are now rich such as the U.S., Japan and Western Europe, of the construction of infrastructure such as railroads occured 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" (Savage, 1966). Others, such as Robert Fogel (1961, 1991), 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 U.S. than pre-existing river networks. And that 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.

The effect of access to transportation infrastructure on economic development is at the end of the day, an empirical question. And the main difficulty in answering this question is reverse causality. Does infrastructure cause economic development? Or does economic development increase demand for infrastructure? This corresponds to the following thought experiment: if a government were to "randomly" place transportation infrastructure throughout its country, will the affected regions grow more than otherwise?

A number of recent papers have used carefully constructed estimation strategies to deal with this endogeneity issues. Michaels (2007) looks at the effect of highway construction in the U.S. 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 (2008) studies the effects of railroad construction in 19th century India using a difference-in-difference approach. And Keller and Shue (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 where 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 as predicted by trade theory.[1]

This paper differs from this literature in two ways. First, to deal with endogeneity issue of where the transportation infrastructure is constructed, we construct a variable indicating the distance to straight lines joining historical cities in China and city ports. The idea is that being on or near the straight line between two major cities makes it much more likely that a transportation route will be built that connects that area to the rest of the world, compared to a similar area off of the straight line. We show that the distance from the straight line is a good predictor of distance to railroads, the one form of transportation infrastructure for which we have detailed data. We emphasize however that the distance from the straight line joining two major cities is not, however, a good instrument for railroad to the exclusion of other forms of transformation, since it is likely that other communication roads have historically been present on those lines.

Second, we focus on the causal impact of access to a transportation route on per capita GDP and per capita GDP growth across counties in China in the period 1986-2003. 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.

A related paper is Atack et al. (2009), who focus on the effect of railroads on urbanization and population growth in the U.S. 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. They find a strong effect on urbanization but a small effect on population growth.

Using county level data reported in all the provincial Statistical Yearbooks that were ever published during 1984-2003 and distance variables constructed using ArcGIS, we show that proximity to the straight line joining important historical cities has a positive and significant effect on GDP growth: counties that are 1% closer to a straight line experience growth rate that are 0.019% higher. We do not find a significant effect on GDP levels, population, or the composition of population. To illustrate the magnitudes, we imagine that the entire effect comes from the presence of a railroad near the line and estimate the effect of railroads using

[1] A related study is by Demurger (2001). She investigates the contribution of infrastructure in China's economic growth by estimating a growth model with provincial level panel data. She finds that transport facilities and telecommunications play significant roles.

2SLS. On average, increasing distance from railroads by 1% decreases annual GDP growth by 0.12-0.28% across sectors.

In addition, we also investigate whether the main results reflect business "displacement" or an effect on aggregate production. Since we are comparing counties that are closer to the straight line with counties that are further, it may be that the effect we detect is driven by businesses (and other economic activities) that would have otherwise been located further away from the line. Moving themselves to be nearer the line to take advantage of the transportation infrastructure. In the extreme case, this means that our estimated benefits of proximity to railroads reflect no overall gains in production or growth, and is only capturing redistribution across space.[2] This is, of course, a problem with any comparison of connected and unconnected areas: for example it could be that the urbanization that Atack et al. (2009) find was accompanied by de-urbanization elsewhere, and even the price stabilization effect might have come with price destabilization elsewhere, as the traders now focus on the connected areas. There is no direct way of correcting for this problem. We address it indirectly by omitting the counties that are closest to the railroad. If the main effects are driven by displacement, then the estimated benefits of proximity to railroads should become smaller in magnitude if we omit the counties closest to the railroads where businesses presumably located to. We find that this is not the case: estimates remain stable when the counties closest to the lines are excluded from the regressions.

There are several important caveats to interpreting the results. First, our data shows where the output is reported rather than where it is actually produced. It could be that firms relocated their headquarters nearer to the transportation corridor and they report production that occurred elsewhere at the headquarters' location. However, our finding that the effects are similar between the primary sector, which comprises mostly of agriculture, and the other sectors, suggests that this is highly unlikely.

Second, as already indicated, our estimates should be interpreted as the effect of being on a transportation corridor, where railways, roads, canals, gas lines and electricity grids may all play a role. Moreover it is the effect of having been on the corridor for tens, or, in some cases, even hundreds of years. In other words, if, in the likely case that transportation creates economic opportunities and economic opportunities then creates need for further transportation infrastructure, and so on, the effect we measure is the cumulative effect of all of those processes. This is the price we pay for trying to look at the effect on overall economic activity, which is something that is only measured in recent times.[3]

[2] The distributional effects of infrastructure have been found by Duflo and Pande (2005) who found that dams in India decrease poverty in downstream districts but increase poverty in regions where they are built.

[3] It is less of a problem with the studies we mention above that measure the impact of some form of infrastructure investment in the immediate aftermath of its construction.

These caveats not withstanding, our results suggest a policy which "randomly" places transportation infrastructure will have a positive economic effect on those areas. This is as far as we can go. We cannot use our results to estimate the social or private return on investing in transportation infrastructure because we have no idea of the relevant costs.[4]

The paper is organized as follows. Section 2 discusses reasons why transportation matters. Section 3 provides the background and the empirical strategy. Section 4 describes the data. Section 5 presents the results. Section 6 offers concluding remarks.

2 Why does transportation infrastructure matter?

There are a number of reasons why good transportation infrastructure is advantageous for economic development. First, it reduces trade costs and promotes market integration. This should reduce price volatility and reallocate resources in line with 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 come back easily whenever needed. It is easier to lend to someone whose project you can visit. It is easier to put 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. Fourth, and more intangibly, the freer movement of people and goods may bring with it new aspirations, new ideas, and information about new technologies.

All of these reasons can generate increases in output, which, in the short run, also leads to faster growth. In China, the corridors themselves are typically more than a hundred years old. If China were an economy nearing steady state, one might have imagined that the growth impact would be small. But the years we study (1986-2003) are years of explosive growth in China. Therefore, all of China is probably best thought of as being in transition. Given this, it is not hard to imagine that there would be both level and growth effects.

The fact that the increased use of transportation infrastructure is important to the growth process does not mean that public investment in roads will promote growth. It may be that the infrastructure that gets created by governments in the absence of strong demand pressures is entirely useless. But this is what in effect happened to the part of the Chinese countryside that happened to be on the way from one important city to another. Because the two cities desired to be connected, all the places along the way got connected as well, despite there being no special demand pressure in these places. The purpose of this study is

[4]See Fleisher et al. (2009) for a cost-benefit analysis of hypothetical investments in human capital and infrastructure.

to understand whether these locations that happened to be near a transportation route by historical "accidents" benefited.

Why would we imagine that investment in transportation infrastructure in more or less arbitrary locations would promote growth? Wouldn't private investors have put in the money if it were really profitable? Unfortunately, as mentioned above, we cannot estimate the return in these infrastructure projects and therefore cannot rule out the possibility that despite the growth, these were money-losing investments (both from the private and the social point of view). However, putting aside this possibility, it is entirely plausible that even profitable investments in transportation infrastructure would not get taken up by the market. In part, this is because while most forms of transportation infrastructure are excludable in principle, excluding people is expensive. A large number of people used to ride the subway in Boston without paying (and some still do) and it took a substantial investment in technology to stop it. Even if detection were perfect, the idea of stopping to pay a toll every time you take a turn is clearly ridiculous. Hence, it is not surprising, that everywhere in the world, local roads have free access.

Even if excludability were not a problem, one must consider the divergence of private and social returns. The free flow of ideas and technologies along the corridor is clearly impossible to price. It is also not clear that parents internalize enough of the value of their children's human capital to guarantee that they would be willing to pay for roads so that their children can go to school.

Of course, the market is not the only possible supplier. The forces of local political economy, as emphasized by Foster and Rosenzweig (2003) also respond to demand for infrastructure. Roads (and to a lesser extent, railroads) can be thought of as a club good. Under certain conditions, clubs can efficiently supply local public goods. However, these conditions are stringent. All the reasons of why the market demand for infrastructure may be too low also apply here. Moreover, the fact that each piece of road or railroad is useless without complementary investments in neighboring areas creates a formidable coordination problem.

Public investment in infrastructure may therefore be desirable, though as we have already noted, we would need cost data to be able to speak definitively about that. With the data at hand we can only investigate whether it would promote growth.

3 Empirical Strategy and Historical Background

As explained above, the basic idea behind our empirical strategy is to examine the correlation between the distance to the nearest straight line connecting two historical cities and the outcomes of interest. The way we implement it is to start with the set of important historical cities in China *circa* 1860. 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. This is because most of the early Chinese railway lines were first put into place by non-Chinese powers during the end of the 19th and beginning of the 20th Century to connect the Treaty Ports to important Chinese cities at the time. This allowed fast troop deployment for the League of Eight Nations, who were restricted to station their troops in the Treaty Ports, in case of a Chinese "rebellion".

Given this set, 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 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. They are illustrated in Figure 1.

Perhaps not surprisingly, the lines drawn this way coincide well with railroads constructed during the late 19th century.[5] They do not coincide well with the railroads in North Western China (Xinjiang province) and Tibet, where construction occurred under the Communist government after the 1970s, or the railroads in North Eastern China (Manchuria), where construction was mainly done by Czarist Russia around the turn of the century, independent warlords, and a *de facto* colonial Japanese government during the 1920-30s. Our main sample will exclude Xinjiang, Tibet, Inner Mongolia and Manchuria because the recent constructions had complex political aims while the construction in the 1920s and 1930s reflected competitive railroad building as part of an attempt to define political authority. (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 are 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 that the line does indeed proxy for transportation infrastructure, we estimate the correlation between distance to the line and distance to railroads. Note that the prevalence of paved motorways mean that there is little variation in measured access to roads across counties in our data today (though there may have been historically). 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.

[5] 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.

$$\ln dist\_rr_{cpt} = \delta \ln dist\_line_{cp} + \rho_p + \gamma_t + \varepsilon_{cpt} \tag{1}$$

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. As noted earlier, this can in general not 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.

$$Y_{cpt} = \beta \ln dist\_line_{cp} + \alpha \ln dist\_seg\_term\_city_{cp} + \rho_p + \gamma_t + \varepsilon_{cpt} \tag{2}$$

The outcome for county c, province p and year t, Y_{cpt}, is a function of: the natural logarithm of shortest distance to the line for county c in province p, $dist\_line_{cp}$; the natural logarithm of the nearest distance to the urban centers which form the terminals of line segments, $dist\_seg\_term\_city$; province fixed effects, ρ_p; and year fixed effects, γ_t. One concern with our strategy might be that the distance from line variable might be picking distance from big city at the origin of the line. To address this, we control for distance to the large cities that make up the ends of the line segments. Standard errors are clustered at the county level. If proximity to the line is beneficial, then $\hat{\beta} < 0$.

Purely for the sake of comparison we also estimate a structural equation that in effect assigns the entire value of being in transportation corridor to railroads both using OLS and 2SLS, in the latter case using the distance to the line as an instrument:

$$Y_{cpt} = \beta \ln dist\_rr_{cp} + \alpha \ln dist\_seg\_term\_city_{cp} + \rho_p + \gamma_t + \varepsilon_{cpt} \tag{3}$$

4 Data

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 data is then matched at the county level to data from the Provincial Statistical Yearbooks from China from 1986-2005. We collected data from all published yearbooks that reported county-level statistics on GDP.[6] The variables which were consistently reported included GDP

[6] See References for a complete list. Hard copies of these books are stored in the National Library in Beijing. Research assistants scanned and entered the data into computer format.

and population. 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.

Finally, in terms of where we measured outcomes, we excluded the former Treaty Ports that are now province-level municipalities (e.g. Shanghai, Tianjin, Chongqing) and other large cities that are on the segment ends 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.

Our final sample 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 using sample comprises of an unbalanced panel with 3,039 county-year level observations.

Table 1 describes the data. On average, counties are 38 km from the nearest railway, 67 km from the nearest point on lines connecting Treaty Ports to historical cities (our constructed instrument), 70 km from the nearest point on lines connecting segment cities, 377 km away from the nearest historical city, 634 km away from the nearest Treaty Port, and 138 km from the nearest urban center with more than five million individuals. On average, 424,655 individuals reside in each county. This is similar to the national average. Almost 90% are agricultural. This is higher than the national average of approximately 74% due to our exclusion of large urban centers that are at the end of rail segments. GDP is deflated using a national level CPI. Disaggregated region level CPIs are not available over time. Per capita GDP is approximately 3,240 RMB, slightly higher than the national average during the period. Most of this is from the primary sector which includes agriculture and mining, and the secondary sectors which covers manufacturing. Tertiary sectors, i.e. service industries, contribute relatively little to GDP per capita in this period. Per capita GDP is growing at approximately 22% annually. This is higher than the national average during the period which was approximately 12%.

Next, we split the sample into ten equal frequency groups according to distance from the line. We report means for selected variables for these groups in Table 2. Column (1) shows that there is much more geographic dispersion in the 20% of counties furthest away from the line. Column (2) shows that distance to the constructed line is highly correlated with distance to railroads, suggesting that the former is a good proxy for transportation network. Column (3) shows that GDP typically decreases as distance increases. Column (4) shows that total population is also approximately decreasing with distance. Column (5) shows that the agriculture population, as a fraction of total population, is increasing with distance.

5 Results

5.1 Lines, Railroads and Transportation Networks

Table 2 shows the estimates of the correlation between the distance to the nearest railroad and the 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 (1). Distance is measured in terms of kilometers. We estimate the correlation for the full sample, a restricted sample where the nearest 10% of counties are omitted, and a restricted sample where the nearest 20% counties are omitted. For all samples, the effects are statistically significant at the 1% level with and without the control for distance to segment terminal cities. The positive and significant estimates for distance to the nearest big city reflects the fact that all big cities have railroads. Hence, being closer to a big city increases proximity to railroads. The coefficient increases in magnitude as we omit the nearest counties. This reflects the fact that there is more variation in distance as we go further away.

5.2 The Effect of Distance from the Line

Next, we estimate the reduced form effect of the distance to the line on the outcomes of interest. The estimating equation is identical to the previous estimating equation, equation (1), except that the dependent variables are now the outcomes of interest. The estimates for the full sample are shown in Table 4. Columns (1)-(4) show that the correlation between the distance to the line and GDP levels are not statistically significant. However, Columns (5) -(8) show that distance from these lines are negatively correlated with GDP growth in aggregate and across sectors. The estimates are statistically significant at the 10% levels or better. They suggest that increasing the distance to the railroad by 1% would reduce total GDP growth of a county by 0.019%.

Primary industries are mainly comprised of household level agricultural production and these households would not report their revenues as part of a larger firm. This means that since the effect of being closer to a communication line is as large for the primary sector as for other sectors, the estimated effect of railroads is not likely to reflect the fact 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:

To investigate whether the effect we observe in the full sample is caused by displacement, we repeat the estimation on a sample where the 10% nearest counties are excluded, and then again on samples where the nearest 20% are excluded. If the full sample results are caused by productive firms relocating to be 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 A and B show that this is not the case. Our estimated effects on GDP growth are similar in magnitude across subsamples. The effect on levels remain insignificant, while, for the most part, the coefficients on growth remain significant.

Finally, Table 6 examines the effect of distance on population, population growth, and the share of the population in agriculture. None of these variables appear to be affected by the distance to the road. This is consistent with the restriction in migration in contemporary China. However, since these communication lines are ancient, this is a somewhat surprising result. Nevertheless, it reinforces our previous conclusion that the results are more likely to be due to net growth, rather than to displacement or differential reporting of activities. We believe that the effects are unlikely to have been driven by strategic behavior since there appears to have been no migration or change in the composition of the labor force of the composition of the GDP in response to the railroads.

5.3 Estimating the effects of Railroads: OLS and 2SLS

To what extent did railroads locate in places that were already richer or growing fast? To get a sense for this, we compare the OLS estimate of the correlation between railroad and GDP levels and growth to an IV estimate using the distance to the straight line as an instrument.

Table 7 Panel A shows the OLS estimates from equation (3). Columns (1)-(4) show that increasing distance to railroads by 1% is associated with a decline in per capita GDP levels of 0.08-0.27%. The estimates are all statistically significant at the 1% level. Columns (5)-(8) show that there is no correlation between per capita GDP growth and proximity to a railroad.

The 2SLS estimates are shown in Panel B. The estimates are larger in magnitude and much noisier than the OLS estimates, both in growth and in levels. Like the reduced form estimate of the impact of the distance to the straight lines, the IV estimate of the impact of railroad on growth are significant, and the estimate of the impact on levels are insignificant. The point estimates in columns (5)-(8) suggest that increasing distance to railroads by 1% decreases per capita GDP growth by 0.12-0.18%.

Recall that the IV estimates of the impact of railroads are upper-bounds of the causal effect of railroad because they also capture the effects of other communication channels. The finding that the IV estimates are larger and more imprecise than the OLS estimates (both for growth and level) means that we cannot conclude that railroads are not optimally placed in richer or faster growing places.

6 Conclusion

These results are encouraging for those who believe that investment in transportation infrastructure can promote growth. But as already noted, this does not give us the return on such investment and therefore we cannot say whether this is where we ought to invest. Finding credible ways to estimate or even bound the social returns remains a very important next step in this research agenda.

References

[1] Jeremy Atack, Fred Bateman, Michael Haines and Robert A. Margo (2009) "Did Railroads Induce or Follow Economic Growth? Urbanization And Population Growth In The American Midwest, 1850-60" NBER Working Paper 14640.

[2] Banerjee, Abhijit, Esther Duflo and Nancy Qian (2006) "The Railroads to Success: Evidence from China". MIT Working Paper.

[3] Demurger, Sylvie (2001) "Infrastructure Development and Economic Growth: An Explanation for Regional Disparities in China?" *Journal of Comparative Economics*, 29, pp. 95-117.

[4] Donaldson, David (2008) "Railroads and the Raj: the economic impact of transportation infrastructure" LSE Working Paper.

[5] Duflo, Esther and Pande, Rohini "Dams", *Quarterly Journal of Economics*, 122(2):601-646.

[6] Fleisher, Belton, Haizheng Li and Ming Qian Zhao (2009) "Human capital, economic grwoth, and regional inequality in China" *The Journal of Development Economics.*

[7] Fogel, Robert (1962) "A Quantitative Approach to the Study of Railroads in American Economic Growth: A Report of Some Preliminary Findings" *The Journal of Economic History*, Vol. 22, No. 2, pp. 163-197.

[8] Fogel, Robert (1964) *Railroads and American Economic Growth: Essays in Econometric History.* Baltimore: Johns Hopkins Press.

[9] Keller, Wolfgang and Shue, Carol "Institutions, Technology, and Trade" University of Colorado Working Paper.

[10] Michaels, Guy (2008) The Effect of Trade on the Demand for Skill - Evidence from the Interstate Highway System" *Review of Economics and Statistics*, 90(4), November.

[11] Savage, Christopher (1966) *An Economic History of Transport.* Hutchinson: London.

Figure 1: Lines Connecting Treaty Ports (Original Four) and Historical Cities

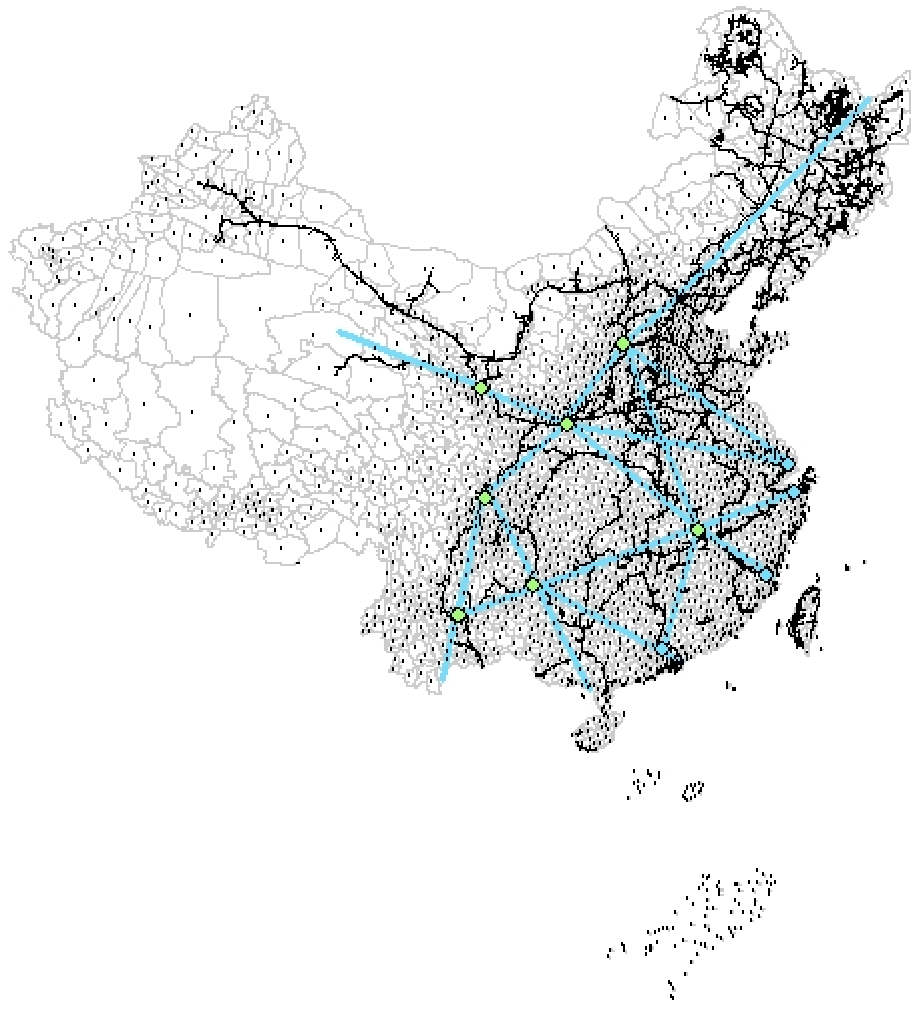

Table 1: Descriptive Statistics

Variable	Obs	Mean	Std. Err.
A. Distance (Km)			
Railroad	3605	38.02	0.63
Treaty Port Lines	3605	67.31	1.02
Big City Lines	3605	78.51	1.22
Historical City	3605	377.38	2.81
Treaty Port	3605	634.09	7.36
Big City	3605	137.63	1.35
B. Population (Individuals)			
Total	3039	424655	5416
Agriculture	994	306675	7557
Agriculture Fraction	994	0.89	0.03
C. GDP (RMB)			
GDP PC	3039	3240	808
GDP PC Primary	2526	1251	337
GDP PC Secondary	2526	1157	374
GDP PC Tertiary	2435	785	199
D. GDP Growth (Annual %)			
GDP PC	2602	0.22	0.03
GDP PC Primary	2089	0.23	0.04
GDP PC Secondary	2089	0.29	0.04
GDP PC Tertiary	1975	0.17	0.03

Table 2: Descriptive Statistics -Population and GDP by Distance to Line

Distance to RR by Quantile	Obs	(1) Distance to Line Mean	(1) Distance to Line Std. Dev.	(2) Distance to RR Mean	(2) Distance to RR Std. Dev.	(3) GDP (Mil RMB) Mean	(3) GDP (Mil RMB) Std. Dev.	(4) Tot Population Mean	(4) Tot Population Std. Dev.	(5) Fraction of Agric Mean	(5) Fraction of Agric Std. Dev.
1	286	5.77	3.23	19.44	20.03	1054.00	760.50	422293	283163	0.98	1.12
2	285	15.31	2.94	23.98	23.22	954.10	873.10	424209	293640	0.85	0.30
3	289	25.35	3.18	29.03	25.45	1155.00	1019.00	475983	372911	0.80	0.29
4	290	35.36	3.66	20.85	16.31	982.10	844.00	422704	282798	0.81	0.34
5	280	47.70	3.90	38.30	35.22	976.80	741.80	425896	322134	0.97	0.93
6	289	62.04	4.99	39.31	33.02	912.00	705.80	448660	314260	0.85	0.28
7	294	74.36	3.14	35.50	25.17	1274.00	1038.00	437407	257186	1.16	1.69
8	286	89.49	6.99	40.64	34.64	1128.00	904.00	448541	297763	0.86	0.30
9	286	120.79	12.63	49.60	40.34	912.10	1137.00	413206	269699	1.11	1.21
10	292	199.92	42.40	58.47	46.98	668.70	470.80	453884	321808	0.77	0.41
Total	2877	35.61	33.59	67.90	57.32	1002.00	882.60	437535	303685	0.91	0.82

Table 3: OLS Estimates of the Correlation between Distance to the Constructed Lines and Distance to Railroads

	Dependent Variable: Ln Distance to RR					
	Full Sample		Omit Nearest 10%		Omit Nearest 20%	
	(1)	(2)	(3)	(4)	(5)	(6)
Ln Distance to Treaty Lines	0.212	0.166	0.248	0.196	0.334	0.272
	(0.069)	(0.068)	(0.118)	(0.117)	(0.144)	(0.147)
Ln Distance to Segment Terminal Cities		0.367		0.368		0.333
		(0.104)		(0.126)		(0.135)
Observations	2877	2877	2592	2592	2315	2315

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

Table 4: The Correlation between the Distance to Constructed Lines and Per Capita GDP

	Dependent Variables							
	Ln GDP				GDP Annual Growth			
	All	Primary	Secondary	Tertiary	All	Primary	Secondary	Tertiary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln Distance to Line	-0.047	-0.055	-0.057	-0.046	-0.019	-0.028	-0.027	-0.025
	(0.046)	(0.048)	(0.061)	(0.052)	(0.010)	(0.011)	(0.012)	(0.010)
Observations	2877	2364	2364	2273	1566	1386	1386	1343

All regressions control for ln distance to segment terminal cities, year fixed effects and province fixed effects. Growth regressions (5)-(8) control for two and three years of lagged GDP per capita. Standard errors are clustered at the county level.

Table 5: The Correlation between the Distance to Constructed Lines and Per Capita GDP Omitting Nearby Counties

	Dependent Variables							
	Ln GDP				GDP Annual Growth			
	All	Primary	Secondary	Tertiary	All	Primary	Secondary	Tertiary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Omit 10% Nearest Counties								
Ln Distance to Line	-0.021	-0.031	-0.018	-0.013	-0.021	-0.025	-0.036	-0.028
	(0.058)	(0.066)	(0.086)	(0.062)	(0.013)	(0.015)	(0.018)	(0.012)
Observations	2592	2113	2113	2031	1422	1246	1246	1206
B. Omit 20% Nearest Counties								
Ln Distance to Line	-0.050	-0.068	-0.085	-0.025	-0.019	-0.029	-0.038	-0.021
	(0.070)	(0.075)	(0.103)	(0.075)	(0.018)	(0.021)	(0.024)	(0.014)
Observations	2315	1882	1882	1804	1272	1108	1108	1068

All regressions control for ln distance to segment terminal cities, year fixed effects and province fixed effects. Growth regressions (5)-(8) control for two and three years of lagged GDP per capita. Standard errors are clustered at the county level.

Table 6: The Effect of Distance from the Constructed Lines on Population

	Dependent Variable					
	Levels			Growth		
	LnTotPop	LnAgPop	Ag Frac	LnTotPop	LnAgPop	Ag Frac
	(1)	(2)	(3)	(4)	(5)	(6)
Ln Distance Line	-0.046	-0.177	0.022	-0.002	-0.006	0.009
	(0.070)	(0.154)	(0.045)	(0.004)	(0.014)	(0.009)
Observations	2306	740	740	1975	596	596

All regressions control for ln distance to segment terminal cities, year fixed effects, and province fixed effects. Standard errors are clustered at the county level.

Table 7: OLS and 2SLS Estimates of the Effect of Distance from Railroad

	Dependent Variables							
	Ln GDP				GDP Annual Growth			
	All	Primary	Secondary	Tertiary	All	Primary	Secondary	Tertiary
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. OLS								
Ln Distance to RR	-0.155	-0.076	-0.287	-0.154	0.003	0.001	-0.008	0.002
	(0.038)	(0.037)	(0.056)	(0.041)	(0.009)	(0.009)	(0.010)	(0.009)
Observations	2877	2364	2364	2273	1566	1386	1386	1343
B. 2SLS								
Ln Distance to RR	-0.285	-0.313	-0.328	-0.303	-0.124	-0.182	-0.175	-0.160
	(0.293)	(0.302)	(0.346)	(0.364)	(0.086)	(0.119)	(0.117)	(0.105)
Observations	2877	2364	2364	2273	1566	1386	1386	1343

All regressions control for ln distance to segment terminal cities, year fixed effects and province fixed effects. Growth regressions (5)-(8) control for two and three years of lagged GDP per capita. Standard errors are clustered at the county level.