

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

Abhijit Banerjee,[†] Esther Duflo[‡] and Nancy Qian[§]

Revised Version: February 2012

Abstract

This paper estimates the effect of access to transportation networks on regional economic outcomes in China over a twenty-period of rapid income growth. It addresses the problem of the 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 levels across sectors, but no effect on per capita GDP growth. We provide a simple theoretical framework with empirically testable predictions to interpret our results. We argue that our results are consistent with factor mobility playing an important role in determining the economic benefits of infrastructure development.

*This paper updates and supersedes “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 Naomi Lamoreaux, Tom Rawski, Thomas Piketty and the participants at the 2004 MacArthur Network for Inequality Conference in Beijing, the China Summer Institute, and the 2011 IGC Conference in London for very helpful comments. We thank Zhichao Wei, Gongwen Xu and the large team they assembled to collect the data; and Giovanni Zambotti and Ceren Baysan for invaluable assistance with GIS. We acknowledge financial support from the IGC.

[†]MIT, J-PAL, BREAD, NBER, CEPR, banerjee@mit.edu

[‡]MIT, J-PAL, BREAD, NBER, CEPR, eduflo@mit.edu

[§]Yale University, BREAD, NBER, CEPR, nancy.qian@yale.edu

“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

1 Introduction

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 that the historical construction of infrastructure such as railroads coincided with periods of rapid economic growth in Western Europe, Japan and the United States. Today, it is indisputable that richer countries have dramatically better transportation infrastructure than poorer ones. However, policymakers considering the trade-offs of investing in infrastructure must consider several related questions. First, they must consider the question of causality: is infrastructure development a worthwhile object of policy, or is it better to rely on the natural forces of the market and/or competition between local jurisdictions to endogenously provide the necessary infrastructure when the demand is there? For example, Fogel (1962, 1964) famously argues that one of the most frequently mentioned historical innovations in transportation infrastructure, railroads, was less effective for economic development in the United States than the pre-existing river networks and that this misdirected investment was a result of government policies for promoting railroads.

Second, policymakers are typically concerned about the distributional effects of infrastructure, which are by no means obvious. On the one hand, for fixed factor endowments, the increased access to markets and ideas should benefit all regions. For example, in the historical context of the United States, it has been argued that transportation infrastructure gives rise to more cities, which then turned into “engines” of growth for the country as a whole.¹

¹For example, see the World Bank’s World Development Report 2009 on *Reshaping Economic Geography*

On the other hand, transportation infrastructure increases the access of rural regions to cities, and the well-known agglomeration effects of cities may cause productive capital and skilled labor to move from rural regions to cities over time, with the result that those who remain in rural areas receive very limited benefits from urbanization or even become impoverished. Along similar lines, it has been argued that the expansion of motor road networks in the United States promoted large-scale suburbanization and left many cities without a viable economic model (Glaeser and Gottlieb, 1999).

This paper makes progress in understanding the impact of access to transportation infrastructure by examining the causal effect of access on economic performance in different regions in China during a twenty-year period of rapid growth. We ask the straightforward question: Do areas that are “quasi-randomly” assigned to have better access to transportation networks consequently have better economic outcomes in the long run? Specifically, we attempt to empirically examine two closely related questions. First, we ask whether access to better transportation enriches the *average region* that is affected (because it draws in or generates more new economic activities) or impoverishes it (because it becomes easier for human and physical capital to exit). Second, we ask whether areas that have better access to transportation networks benefit much more and serve as engines of growth when new economic opportunities arise and growth becomes possible.

For our discussion, it is important to keep three points in mind. First, our focus is on long term effects. We are interested not just in the impact on trade and prices that result from greater access, which tends to be relatively short-term, but also in the subsequent changes in the patterns of localization of economic activity as people and factories relocate. Second, the emphasis on understanding the effect of infrastructure for the *average* location is crucial to our study since it is entirely possible that some of the largest cities benefit from infrastructure through greater concentration of resources while most other places lose out. Finally, there cannot be one definitive answer to these questions, since the answer will clearly depend on the starting point – i.e. the first road to connect the agricultural hinterland to a port is very different from the fifth such road.

by Aoyama and Horner (2009) for a nuanced statement of this view.

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 and a set of Western Colonial powers built railroads connecting the historical cities of China to each other and to the newly constructed so-called “Treaty Ports”.² We identify our average “treated” areas to be those that were close to the straight line connecting the same set of cities. Our analysis excludes the termini cities, where there are obviously additional termini effects. Our strategy compares areas closer to the lines to areas further away and interpret the result of this comparison as the overall effect of any transportation infrastructure – the original railroads and any other infrastructure later added – along these historical transportation corridors.

This strategy has a number of advantages. First, it provides us with 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, 1986-2006, coincides 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 1986-2003.

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 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 Chinese Ministry of Agriculture, which shows that distance has no significant effect on household income. This might reflect the effects of the two contending forces that we discussed earlier. For the estimates of the effect of proximity on growth, we find a precisely estimated zero effect. The estimated elasticity between distance to the line and annual per

²For example, see Pong (1973).

capita GDP growth is -0.002 and statistically insignificant (the standard error is 0.003). Places close to the line grew exactly as fast as places further away.

Our finding that better access to transportation networks does not have a large impact on the (relative) economic performance of those areas is consistent with the Fogelian view that transportation infrastructure by itself does not really do very much, excepting perhaps where there was already a demand for it. Based on similar logic, China scholars have criticized public investment in transportation infrastructure in China after 1990 (Huang, 2008).

However, in the interest of understanding the potential benefits and limitations of infrastructural investment in developing countries more fully, we also consider an alternative and complimentary interpretation. We ask whether there are any obvious characteristics of China (or developing countries in general) that could cause the *measured* benefits of infrastructure to be small even if better transportation causes substantial gains to GDP. In particular, our model emphasizes the role of factor mobility – i.e., labor is assumed to be completely immobile, while capital is assumed to move at a cost. Using a simple model, we argue that the empirical evidence is consistent with a version of the model where capital is less mobile than goods such that even distant places retain high levels of capital. The result is that the level differences in productivity will tend to be relatively small and both near and far places will be involved in the production of exports. This, in turn, means (in the world of our model) that both locations will gain equally in proportional terms from China's integration into world markets. Therefore, even though better transportation does help China as a whole to gain more from trade, GDP level differences between well- and poorly connected areas can be small and there may be no differences in growth rates between the two areas.

The premise of this model is, as we emphasized above, that goods move more easily than capital. Unfortunately we cannot directly observe the relative mobility of goods and capital. However, our model tells us that relatively low mobility of capital is likely to be associated with a situation, where there is higher inequality in better connected areas, under the assumption that the direction of capital movement is from less connected areas towards better connected ones. Using regional income inequality data computed from the *National*

Fixed Point Survey collected by China’s Ministry of Agriculture (1987-2005), we do find that inequality is higher in better connected areas.

In assessing what general lessons one can learn from our results, one should consider whether our results and theory are driven by conditions specific to the Chinese context. For example, one may be concerned that the marginal effect of access to infrastructure is especially low in China due to the massive investment into infrastructure during recent years on the part of the government.³ We believe that this is highly unlikely to be driving our results because our data shows large variation in access to infrastructure. In our data, the distance to the railroad for the counties that are the nearest the line (defined as the nearest decile) are only a third of the distance for those that are the furthest (defined as the furthest decile). Similarly, the nearest counties have more than twice the length of highway relative to the furthest counties (despite the fact that the latter are almost eight times as large in area).

Another concern for external validity is that the lack of factor mobility stems from the Chinese government’s attempts to control labor mobility and that the empirical findings are not easily generalizable to the context of other developing countries.⁴ We acknowledge that the Chinese government may be unique in implementing an explicit policy for controlling migration for so long. However, it is important to note that the actual patterns of low levels of migration are not unique to China. In particular, the main policy effort has focused on unskilled low-wage rural workers (Meng, 2005), a group that has been found to be relatively immobile in other contexts such as India (Munshi and Rosenzweig, 2009). Moreover, the theory predicts similar effects regardless of which factor (e.g., labor, capital) is immobile and the immobility of capital has been documented in several developing country contexts.⁵

A growing number of recent papers have developed compelling identification strategies to evaluate the impact of transportation infrastructure (e.g., Michaels, 2008; Donaldson, 2010; Keller and Shiue, 2008).⁶ These studies examine the effect of transportation infrastructure

³For example, see Huang (2008) for a discussion on infrastructure investment in recent years in China.

⁴For example, see West and Zhao (2000) for a review of studies on labor migration.

⁵For example, see Duflo (2004) for evidence of limited capital mobility within Indonesia.

⁶Michaels (2008) examines the effect of highway construction in the United States in the 1950s, using both a *difference-in-difference* (DD) approach and an instrumental variables approach, where he exploits the

from the point of view of market integration and the focus is on price convergence and changes in the relative price of factors along the lines predicted by trade models. The results suggest that transportation infrastructure favors greater price convergence and that factor prices shift in the direction predicted by trade theory.

Our study differs from earlier studies in its focus on the longer-run and more macro question: do areas that benefit from access to the reduction in trade costs and perhaps other costs become wealthier as a consequence? This is by no means obvious even if there is clear evidence that trade and other flows such as migration increased when infrastructure became available. 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 et al. (2009), which finds that access to railroads has a strong positive effect on urbanization but a small effect on population growth in the United States. Similarly, in a study contemporaneous to ours, Faber (2009) uses a straight-line instrument to study the impact of recently constructed highways in China on economic performance.⁷

Moreover, our paper provides a potential interpretation of the lack of infrastructure effects which is of some independent interest. The idea that the lack of factor mobility might limit the impact of better infrastructure is of considerable relevance to many developing countries that are currently investing in improving their infrastructure.

The rest of the paper is organized as follows: We start with a brief review of the literature in Section 2. 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

variation in access caused by the fact that highways tended to be built in either a North-South direction or an East-West direction starting from big cities. Donaldson (2010) studies the effects of railroad construction in 19th century India using a DD approach. Keller and Shiue (2008) uses a similar strategy to examine the opening up of railways between regions within Germany.

⁷While Atack et al. (2009) primarily uses a DD approach, it also constructs an instrument for the distance to the railroad based on the straight line between the start and end points of a railway line. Similarly, Faber (2009) both uses a straight-line instrument as well as estimates of transportation costs for exogenous variation in access to transportation. Both studies generously credit the straight-line instrument to an earlier version of our paper (Banerjee et al., 2004).

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 three 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 also 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.⁸

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 secondary and higher education institutions in China are located 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 of primary schools 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.

⁸See, for example, Hu et al. (1997) for an overview of Chinese growth.

For our study, it is important to note the following facts. First, a significant proportion of industrial output in China during our study period came from TVEs. As a percentage of national industrial output, output from TVEs grew from 9% in 1978 to 36% in 1993.⁹ Second, TVE assets are jointly owned by all community residents, which were approximately 400 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.¹⁰ The data show that over half of the profits were re-invested.¹¹ 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 rural areas.

Labor mobility was also restricted. If a worker moved without official permission, she lost access to all public goods. For urban residents, this meant losing 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 to relocate or workers with skills that are needed in specific industries such as construction during the mid- and late- 1990s. But for the rest of the population, permission was extremely difficult to obtain (e.g., Meng, 2005; Meng and Kidd, 1997). Therefore, while the number of migrant workers increased greatly during this period, most of them were temporary migrants who maintained their original residences.¹²

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

⁹See the *Statistical Material of Township and Enterprises*, 1992.

¹⁰See Articles 18 and 32 in *The Regulation on Township and Village Collective Enterprises of the People's Republic of China* (1990).

¹¹See *Statistical Survey of China*, 1992: pp. 67.

¹²There has been numerous studies on migration in China. Zhao (1999) provides a survey of recent evidence.

steadily for the remainder of the reform era. On average, rural real income growth rates 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 lead to a convergence in prices, 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 – e.g., it is easier to migrate to the city if one can return easily whenever needed; easier to lend to a borrower whose project you can visit; and 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 human capital – e.g., 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 imports (e.g. food), and another good which it consumes. These goods could be either identical or differentiated. The key assumption

is that the relative price of the exportable in terms of the importable good 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 $AK^\alpha L^{1-\alpha}(\bar{K})^\beta$ everywhere, where \bar{K} is the average level of K in firms in that region.¹³ 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: $\alpha + \beta < 1$.¹⁴ Assume for the time being that there is no other technology of production.

The key assumption is with respect to factor mobility. We assume that labor does not move: The metropolis has an endowment of labor of L^* while all other regions have an endowment of 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, much of the growth and capital accumulation occurred in rural areas, and it was only later that economic freedoms were extended to urban areas and the urban growth rate crossed its rural counterpart. 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 one is, the more it costs her to send capital to the metropolis. Therefore, she is willing to accept a lower return on capital if it is invested in her own region (e.g., because she can monitor the borrower more easily).¹⁵ We assume that there are no other constraints on mobility (no

¹³We could easily let A vary across the locations to capture differences in the flow of ideas.

¹⁴See Duranton and Puga (2004) for a review of this literature.

¹⁵The equivalent assumption for human capital would be that there is a cost to relocating from one’s home region to the city, but the cost is lower if she is better connected (e.g., because it is easier to travel to and fro).

within-region credit constraints for example).

3.1.2 Analysis of the basic model

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

$$\begin{aligned} w &= p(1-d)A(1-\alpha)\left(\frac{K}{L}\right)^\alpha(\bar{K})^\beta \text{ and} \\ r(1-\rho d) &= p(1-d)A\alpha\left(\frac{L}{K}\right)^{1-\alpha}(\bar{K})^\beta, \end{aligned} \tag{1}$$

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:

$$MK_D + NK_C + K_M = K, \tag{2}$$

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.

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

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

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

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

It follows directly that inequality is higher wherever 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, we compare outputs per worker/capita,

$$y = p(1-d)A\left(\frac{1}{L'}\right)^\alpha (K)^{\alpha+\beta}, \quad (5)$$

which can be written as

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

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 declines with d , but when $\rho > 1$, we might actually observe the reverse, especially when spillovers are large ($1-\alpha-\beta$ is close to zero) and therefore $\frac{\alpha+\beta}{1-\alpha-\beta}$ is large. Once again this is because the better connected region loses more of its capital.

Result 1: In the basic model, output per capita will always be higher and inequality lower in the better connected region as long as capital is more mobile than goods. However when capital is less mobile than goods, the more distant area will have less inequality and the difference in per capita output between the regions will tend to be small and may even be higher in the more distant region.

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

Result 2: In the basic model, the effect of trade opening would be to raise income levels everywhere in proportion and hence have no differential growth effect.

3.1.3 A simple extension

The growth result here is somewhat of an artifact of the way the model is set up. Suppose we add an alternative production technology that uses only labor and produces a perfect substitute for the importable good (e.g. 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 now the wage in the exporting sector, w , needs to be bigger than B for there to be production of exportable goods. In this model, there can be three types of equilibria: type A, where both close and distant locations export; type B, where one of the locations exports and the other does not; and type C, where neither exports. As long as $\rho < 1$, we know that wages, which are proportional to output per capita, will be lower in the more distant location, and therefore, if we are in case B, the distant location will not export. It follows that as long as $\rho < 1$, the effect of trade opening will either be the same in both areas (types A and C), or the more connected area will grow faster.

On the other hand, when $\rho > 1$, it is not clear which of the two locations will have lower wages, and the gap between wages is likely to be small. Therefore we are more likely to be in either type A or C equilibria. Given the high average growth rate of approximately 8% in both close and far regions (see Table 2 column 3), scenario A seems more likely at least for the present. This is consistent with the fact that China now has excellent infrastructure and both near and far places are relatively easily accessed.¹⁶ In any case, in both type A and type C equilibria, the effect of trade opening on growth rates is the same both in close and distant places, unless the effect of the trade shock is just big enough to move one area from not exporting to exporting but not the other.

¹⁶In the model, it is possible for even a very poorly connected area to export because we place no lower bound on the interest rate. But if transportation is really expensive, the interest rate will have to be very negative in the distant areas to permit exporting. It seems likely that capital owners will then prefer to hold cash or gold and therefore, there will not be any exports.

Result 3: In the model with “agriculture”, trade opening is likely to benefit the closer area more in terms of growth rates as long as capital is more mobile than goods, but in the reverse case, growth rates in both close and distant areas should react relatively similarly to trade opening. Moreover, assuming that China is in a type A equilibrium, Result 1 should continue to hold in this case.

To summarize, 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. The lack of a differential growth effect in this scenario is consistent with an overall beneficial effect of transportation infrastructure, which is what allows both close and more distant areas to be exporting.

4 Historical Background and Empirical Strategy

4.1 The Birth of Modern Infrastructure

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. 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 6.4, 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, Xian, 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). These four cities were chosen for their strategic locations. 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. 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.¹⁷ With the exception of Guangzhou, these locations were villages and not prominent historical urban centers prior to becoming Treaty Ports (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.

Moreover, it is important to point out that the Chinese were significantly behind the Europeans in terms of naval technology in 1842, and did not possess a fleet of similar naval gunships for which the Treaty Ports were chosen. More generally, 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 for several centuries leading up to the Opium Wars.¹⁸ Therefore, places such as Shanghai, Ningbo and Fuzhou, while not entirely uninhabited prior to 1842, were only 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.

¹⁷The Treaty Ports were established in *Article 2* of The Treaty of Nanjing, which was signed between 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, Xiamen did not receive significant investment from the West and only became a Treaty Port during the second wave of Treaty Port Relinquishment by the Qing in 1865. Therefore, in our line construction, we omit Xiamen. The other Treaty Ports of the second wave were Tianjin, Niuzhang, Yantai, Zhenjiang, Hankou, Shantou, Taipei and Tainan (Spence, 1990; Pong).

¹⁸See Spence (1990, Ch. 2) for a detailed discussion of China during the 19th Century.

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 that this would facilitate fast British troop deployment from Shanghai and Ningbo to important Chinese cities. 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 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 lines are shown in Figure 1.

As expected, the lines drawn this way coincide well with railroads constructed during the early 20th century.¹⁹ The three places where they do not match well are North-Western China, where construction occurred under the Communist government after the 1970s, partly as an attempt to politically integrate these areas into China; and North Eastern China (Manchuria), where most of the construction was done by a *de facto* colonial Japanese

¹⁹While the railroads suffered much damage during World War II, after the war, the Guomintang (KMT) and then the Communist (post -1949) governments undertook extensive repairs and construction focused on upgrading the physical structure. A comparison of maps from the 1930s to maps from the 1950s indicate that they mostly did not alter the course of the railroads.

government during the 1920-30s as part of their attempt to lay claims to these areas (Figure 1 shows that in Manchuria, most counties have a railroad). For this reason, our estimating sample will exclude Xinjiang, Tibet, Inner Mongolia and the provinces in Manchuria.

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 the *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 lines do indeed proxy for transportation infrastructure, we estimate the correlation between distance to the line and various measures of infrastructure using the following equation:

$$I_{cpt} = \delta \ln L_{cp} + \rho_p + \gamma_t + \varepsilon_{cpt}. \quad (7)$$

Transportation infrastructure in county c in province p and year t , I_{cpt} , is a function of: the natural logarithm of the distance to the nearest line connection treaty ports and historical cities illustrated in Figure 1, L_{cp} ; province fixed effects, ρ_p ; and year fixed effects γ_t . Note that the fact that the line is likely to be corrected with many different types of transportation infrastructure means that L_{cp} is not an excludable instrument for any given infrastructure.

Our main estimating equation is the following:

$$y_{cpt} = \beta \ln L_{cp} + \Gamma Z_{cp} + \rho_p + \gamma_t + \varepsilon_{cpt}. \quad (8)$$

The outcome for county c , province p and year t , Y_{cpt} , is a function of: the natural logarithm of the shortest distance to the line for county c in province p , L_{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 $\hat{\beta} < 0$.

Interpreting β as the causal effect of proximity to the line assumes that the only difference between places near the line and places further away is the distance to the line. This

obviously relies on the terminal cities not being chosen so that the straight line between them would run through economically important regions. This is the reason for why we focus on the ancient cities of China and the Treaty ports – i.e., the historical cities are both sufficiently far from each other and clearly more important than any place between them in the historical era that it is easy to be comfortable with the identification assumption in this context. Similarly, the Treaty Ports were chosen for their suitability for European gunships rather than what laid between them and the historical cities. Note that we restrict our attention to the first four Treaty Ports to avoid the potentially endogenous influences of later Treaty Ports, which may have been chosen for economic reasons (e.g. proximity to economically viable or prosperous regions).

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, an important traditional means for transportation as well as an important input for agriculture (e.g., river beds provide fertile soils). In this case, distance from our line will also capture the distance from the river, which presumably captures many other effects. To address this, our baseline specifications always control for distance to the nearest navigable river.

The baseline estimation also controls for other potentially influential factors, which we will discuss and motivate 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 problems. One issue is that of endogeneity raised earlier. 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, a problem that occurs when there are too many lines. 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. We will return to discuss

this further in Section 6.4.

5 Data

This paper uses data from multiple sources. All maps are obtained in digital format from the Michigan China Data Center. Geographic measures are constructed using ArcGIS software, assuming a Conical Projection. We define centroids of cities and counties. The lines are constructed to connect the centroids of segment cities (using the algorithm described earlier). We compute the nearest distance from each centroid to the straight line, railroads, navigable rivers, the coastline, the country border and segment cities. Figure 1 displays a map of county boundaries, their calculated centroids, our constructed lines, railways and major navigable rivers. Since the road network in China is quite dense, the distance between an arbitrarily defined centroid and the nearest road is not a meaningful measure of access to roads. Our proxy for access to roads is instead the density of roads in each county. Since our regressions will control for county area, we can control for road density by controlling for road length. Therefore, we compute the length of highways (multi-lane highways) and paved roads (paved motor-roads) in each county.

The first outcome measure we examine is county-level per capita GDP. These are from the *Provincial Statistical Yearbooks* from China from 1986-2003 stored in the National Library in Beijing, China. In 2004-6, we collected and digitized data from all published yearbooks that reported county-level statistics on GDP. These data are interesting because they measure production whereas previous studies have mainly focused on prices. However, there are several problems with these data. First, GDP may have been measured using different techniques across provinces and over time. To the extent that these changes are documented or obvious (e.g., changes in the units of measurement), we have corrected for them. But this is clearly still imperfect. 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. There is little documentation on the logic behind the decision of which counties report GDP

and we can do little to correct for it. Our final sample is an unbalanced panel of 310 counties within seventeen provinces.²⁰ In addition to the data on GDP, we collected data on county population so that we can calculate per capita GDP. Figure 2 maps the counties for which we have GDP data.

To address these measurement difficulties, we supplement the analysis with two additional data sets of higher quality. 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 firm-level data from the *Census of Industrial Plants* in 1993 and the *Census of Manufacturing Firms* during 2004-2006. We are able to geocode these data to the county level.²¹ The first survey includes all industrial plants. The second survey samples all manufacturing firms in China that have market capitalization of five million RMBs or more. We will examine two outcomes, the number of firms and their profits. The data are aggregated to the county and year level and form an unbalanced panel of counties. Figure 3 maps the counties for which we have firm data.

The second additional data are village-level data for rural household incomes from the *National Fixed Point Survey* (NFS) for the years 1987-1991, 1993, 1995-2005. There were no surveys in 1992 and 94 for administrative reasons. The NFS is a longitudinal survey of about 320 villages and 24,000 households distributed across all continental Chinese provinces conducted by the research arm of the Ministry of Agriculture (RCRE). The villages were chosen in 1987 to be nationally representative. There is very little attrition. To maintain its representativeness, villages and households are added over time. Therefore, the panel of villages is not perfectly balanced. For this study, we use household level data on income. Each village contains on average 400 households and approximately a third of them are surveyed by the NFS. The large number of households surveyed in each village means that we

²⁰Beijing, Hebei, Jiangsu, Anhui, Fujian, Jiangxi, Shandong, Henan, Hubei, Hunan, Guangdong, Guangxi, Guizhou, Gansu, Qinghai, and Ningxia.

²¹These data are in principle available for other years. However, we only use the 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) and Banerjee et al. (2012). See these studies for a more detailed description of the data.

can examine the within village income distribution.²² Our income variable measures total net income – i.e., the sum of household income (e.g., home production, agricultural production, wages) minus the sum of production costs, excluding labor costs for home production and agriculture. The data are aggregated to the county and year level. The RCRE provided us with income for each decile of the village income distribution and the Gini coefficient for the within village income distribution each year and 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 coefficient. Figure 4 maps the counties for which we have NFS data. Note that the exact location of these villages are confidential. Therefore, our distance variables measure the distance from the centroid of the county that contains the village to the object of interest. This introduces measurement error to the right-hand-side of our estimates for household income that is most likely classical in nature.

For all samples, we exclude the autonomous regions of Tibet, Xinjiang and Inner Mongolia both because these provinces are predominantly non-Han ethnic minorities, 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. The large cities that are on the segment termini are also excluded to avoid the results being driven by the end-points, which are on the line and were chosen because they were important to begin with. It is important to note that other cities on the line (that are not the termini of line segments) are included in our sample so that our estimates will capture any effects that transportation infrastructure may have on the formation or growth of cities.

Table 1 describes the data. Panel A describes the sample with GDP data. On average, these counties are approximately 71 kilometers from the line and 39 kilometers from railroads. The fact that the average distance to railroads is less than the average distance to the line reflects the fact that we constructed many fewer lines than there are railroads to only capture to major transportation networks and to avoid the problem of having too

²²Villages and households are surveyed every year. The survey uses 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. See Martinez-Bravo et al. (2010) for a description of these data.

many lines that we discussed earlier. During our study period, there are very few highways with median dividers in China. On average, a county has only approximately six kilometers of divided highways. Most motor traffic occurred on paved motor roads without dividers. An average county has approximately 84 kilometers of such undivided paved roads. The average county is far from a navigable river, the coastline and the country border.

Note that the data show significant variation in access to transportation infrastructure. This alleviates any concerns that our study cannot detect significant marginal effects of access because high levels of infrastructure investment by the Chinese government causes there to be too little variation in access.

The average population of a county is approximately 201,347. Per capita GDP is 6,834 RMB. The nominal GDP reported in the statistical yearbooks are adjusted by the national CPI. GDP from primary, secondary and tertiary sectors are roughly similar in size. Average per capita GDP growth is 8% in this sample, which is similar to the national average during the study period. Most of the income growth comes from the secondary and tertiary sectors. Note that the number of observations differ across the GDP variables because not every county is engaged in economic activity in every sector.

Panel B displays the descriptive statistics for the sample of household income data. For the sake of brevity, we focus our discussion on the economic variables. The average within-village Gini coefficient is 0.28. The net household income for the median household is on average 5,460 RMB (constant), which is almost twice as much as the income of the 10th percentile household and approximately half of the income of the 90th percentile. Inequality is growing over time. The Gini coefficient increases by 0.001 per year on average. This is driven by higher income growth rate for richer households, although the level of income increases across all parts of the income distribution.

Since there are approximately three people per household in these data, the household income here implies a slightly lower income than the per capita GDP from the sample in panel A. This is not surprising since the earlier sample includes urban and rural areas, while the household income data in panel B only includes rural households, which are on average poorer than urban households. Similarly, income growth is slower in rural areas. Recall

that there are no data for 1992 and 94. Therefore, we interpolate the annual growth rates between 1991 and 93, and 1993 and 95 as the growth rates for each two-year interval divided by two.

Panel C describes the firm data. Again, for the sake of brevity, we focus on the economic variables. On average, there are 82 manufacturing firms in a county. For interest, we divide these firms into three ownership types: firms owned by the state, firms of mixed ownership and firms owned by private individuals. State-owned firms are directly controlled by the state. Mixed-ownership firms are typically privatized state firms for which the state owns most of the equity. Individually owned firms are truly private enterprises that have little connection to the state. The data show that most firms are owned by the state and individuals. There are only a few firms that are owned by a mix of state and private parties. Next, we describe the data on firm profits. These only report profits on counties with at least one firm. Therefore, the number of observations will differ across variables because not every county has a manufacturing firm of a particular type. The high level of reported profits is consistent with the fact that these data sample large firms (more than five million in market capitalization).

Table 2 shows the outcome variables of interest for different distances to the line. These data show that most of the economic measures of interest decline with distance from the line. Most importantly, we do not observe systematic upticks in these measures as we approach the furthest deciles, which is reassuring for the concern that distance from our line bring a count towards another transportation network.

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 the distance to our constructed lines based on equation (7). Distance is measured in terms of kilometers. Panel A shows that distance from the historical lines are positively correlated with distances from railroads, the coastline and segment

cities, and the length of paved roads; negatively correlated with the distance to the country border and the length of highways within a county; and uncorrelated with whether a county is on the coast line or near a navigable river. 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 distances to the segment city controls for the effect of proximity to a large urban terminus. The distances to the nearest navigable river and coastline control for access to traditional methods of transportation that existed before the lines of interest were constructed. The lengths of highways and roads, and total county area, control for the density of motorways that were constructed long after the lines of interest. Controlling for the distance to the country border addresses the possible influences of a “border” effect.²³ Finally, the control for the distance to the coastline also addresses the fact that during the period of our study, economic conditions diverged greatly between the coastal areas and the interior areas. Without this control, one could be concerned that a positive correlation between economic outcomes and distance to our lines is an outcome of faster growth in the coastal areas, which may also be coincidentally closer to our lines on average. In addition to controlling for the log of the linear measure of these distance measures, we also control for the quadratic terms to capture the idea that the costs of distance from transportation may be diminishing over distance (e.g., there maybe increasing returns to profit).

The estimates show that the coefficient for the log distance to the historical line and its standard error are very stable across specifications. The full baseline specification from equation (8) is shown in column (8). It is statistically significant at the 1% level. It shows that the elasticity between the distance to the line and per capita GDP is -0.0672. Note

²³For example, see Feenstra (2002) and the studies referenced there within.

that because the data indicate that the relationship between the distance from historical lines and per capita GDP is log-linear, our main specification in column (8) does not control for the quadratic of the distance from the line.

One way to assess the magnitude of our results is to benchmark our estimates of the effect of distance on GDP across space to the total increase of GDP over time in our sample. In our sample, the 75th-percentile county in terms of distance is 3.8 times further away from line than the 25th-percentile county. Our estimates imply that distance will cause the 75th-percentile county to have almost nineteen percent ($-0.0672 \times 2.8 = -0.188$) lower per capita GDP. During the eighteen years covered by our data, per capita GDP in our sample grew from approximately 2,744 to 9,916 RMB (e.g., the annual growth rate was approximately 7.5 percent), which is approximately a 242% increase. Therefore, a comparison of the effect of distance across space to the increase in GDP over time suggests that the spatial difference attributable to distance from the line is relatively moderate in size.

For the remaining results, we will show only the baseline specification for the sake of brevity. All regressions will control for the full set of baseline controls shown in column (8) of Table 4: the distances to segment cities, the nearest navigable river, the coastline and the country border; the length of highways and paved roads within a county; the total area of the county; the squared terms of each of the aforementioned variables; and province and year fixed effects. As with the results in Table 4, the estimated coefficients for the distance from the line are very similar with different combinations of controls.²⁴

In Table 5, we examine per capita GDP 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. The estimates are statistically significant at the 1% level for per capita GDP in all sectors and tertiary sectors, and at the 10% level for secondary sectors. To interpret the magnitude of the coefficient, consider the estimate in column (1). It shows that the elasticity of per capita GDP with respect to distance from the line is -0.0672. In other words, a 1% increase in distance from the line

²⁴These results are available upon request.

results in a 0.07% reduction in per capita GDP.

Columns (5)-(8) show the estimates of the effect of distance from the line on per capita GDP growth. We calculate per capita GDP growth as the difference between log per capita GDP growth next year and this year for each county, $\ln(\text{pcgdp}_{c,t+1}) - \ln(\text{pcgdp}_{c,t})$. To control for the possibility that poorer regions may experience different rates of growth relative to rich regions for reasons that are independent of access to infrastructure (i.e. income may be mean-reverting), we control for two lags of the level measures of the dependent variable: $\ln(\text{pcgdp}_{c,t-1})$ and $\ln(\text{pcgdp}_{c,t-2})$.²⁵ The estimates are statistically insignificant for all sectors. They are also very small in magnitude, especially when we consider that the mean growth rate in our sample is 4-8% percent per year, depending on the sector. Therefore, we conclude that we find a precisely estimated zero effect the distance from the line on GDP per capita growth.²⁶

Another way to assess the magnitude of the estimates is to make the extreme assumption that being near the line benefits production only through a region's access to railroads. Under this extreme assumption, we can estimate the upper-bound of the effect of the distance from railroads by dividing our main estimates by the estimated correlation between distance to the line and distance to the railroad (e.g., equation (8) with the dependent variable being the log of distance to railroads). This estimate is 0.133 (the standard error is 0.0628), which

²⁵To check that our results are not driven by the particular lag structure of the controls, we alternatively control for 3, 4 or 5 year moving averages of lag per capita GDP. Our results are robust and we find no effect of distance to the line on growth. The estimated coefficients are similarly small in magnitude and statistically insignificant. The sample size becomes smaller as we introduce longer lags and the estimates become more imprecise. These estimates are not shown for the sake of brevity, but are available upon request. Note that one could alternatively control for per capita GDP in the first year of the panel. We do not do this because the unbalanced nature of our panel means that we would lose too many observations.

²⁶Note that the estimates above avoid the Nickell (1981) bias as we do not control for lag growth. To check that our results are not driven by this choice of specification, we also estimate the growth regression using the more traditional method of the *Arellano-Bond System Dynamic Panel Estimation*, where we control for the lag of per capita growth rate. The lags of the level measures are now used as instruments for lag growth rates rather than controls in the second stage equation. The system GMM estimator uses the levels equation from the standard Arellano-Bond estimate to obtain a system of two equations: one differenced and one in levels. The variables in levels in the second equation are instrumented with their own first differences. This method has the advantage over the standard Arellano-Bond Estimate that it typically increases efficiency. Like the standard estimate, it assumes that the lags of the level measures of the growth variable are exogenous to the error term. And it has the additional assumption that the first-differenced instruments used for the variables in levels are also uncorrelated with the unobserved county effects. For references, see Arellano and Bond (1991) and Holtz-Eakin et al. (1988) for more discussion. The estimates are presented in Appendix Table A.1 panel A. The estimates are small in magnitude and statistically insignificant. Thus, they are consistent with our main estimates that distance from the lines have little effect on growth.

means that conditional on all of the baseline controls, doubling a county's distance from the line increases the distance to the nearest railroad by approximately thirteen percent. Dividing the estimate in Table 5 panel A column (1) by 0.133, we calculate that the maximum elasticity of per capita GDP with respect to the distance to the line is 0.5 for all GDP. Dividing the estimate in column (5) by 0.133, we calculate that the maximum elasticity of growth with respect to distance is -0.0156. As pointed out earlier, the distance to the line is not an excludable instrument for the distance to the railroads because it is also correlated with other forms of transportation infrastructure. However, by interpreting these two-stage calculations as the upper-bound effects of railroads, we can starkly illustrate the small magnitude of the effect (e.g. zero effect) of access to transportation on per capita growth relative to the effect on the level differences in per capita GDP.

One potential issue for interpreting for our finding that per capita GDP levels are higher in regions near the line is the possibility of displacement. For example, the placement of transportation may cause a "crowding-in" effect such that firms re-locate to be near the line. This could cause proximity to the line to be positively correlated with production even if the investment in having a line does not increase aggregate (provincial or national) production from when there is no line. 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 choose to relocate to be close to the line to relocate as close as possible to the line).

Table 5 panels B and C provide little support for the crowding-in hypotheses. For example, a comparison of the estimates in columns (1)-(4) between the full sample estimates in panel A to panel C, where the 20% nearest counties are omitted shows that per capita GDP is, if anything, slightly larger in magnitude as we move further away from the line. As with the full sample, we find no effect on per capita GDP growth.

6.3 The Effect on Firms Placement and Household Income

Table 6 shows the estimated effects of the distance from the line on the number and average profits of manufacturing firms. Panel A shows the estimates for the full sample. Columns (1)-(4) show that distance from the line results in fewer firms. The estimates are statistically significant at the 1% level for all firm ownership types. The coefficient in column (1) indicates that increasing the distance by 1% will result in a 0.08% reduction in the number of firms. In columns (5)-(8), we examine log average firm profits. The estimates show that amongst counties that have at least one firm (of the relevant type), distance results in lower profits. The estimates are statistically significant for all firms and publicly owned firms at the 1% and 5% levels. Column (5) shows that a 1% increase in distance results in a 0.1% reduction in average firm profits.

As with our earlier exercise, we assess the magnitude of our estimates by comparing our estimates of the effect of distance on the number of firms across space to the total increase in the number of firms over time. Since the 75th-percentile county in terms of distance from the line is approximately 4.68 times further away than the 25th-percentile county, our estimate in column (1) implies that it should have approximately thirty percent fewer firms ($-0.084 \times 3.68 = -0.309$). During the three years for which our data uses a consistent sample frame (2004-06), the average number of firms per county grew by twelve percent from approximately 83 to 93 firms per county.²⁷ Relative to the change over time, our cross-sectional estimate implies a high elasticity between distance and the number of firms. However, this is partly an artifact of the short time horizon the firm panel data. For example, if the number of firms had grown at the same rate (approximately five percent per year) for eighteen years (which is the sample length of our GDP data), then the number of firms would have grown by approximately 130% from approximately forty to 91 firms per county. Relative to this cumulative growth over the longer time horizon, the implied elasticity between the number of firms and distance to the line appear more moderate.

Repeating the same calculation for average firm profits, the estimate in column (5)

²⁷Recall that the 1993 firm data is from a census of all industrial plants and has a different sampling frame relative to the *Census of Manufacturing* firms which samples all firms with 5 million RMB or more in market capitalization.

implies that firms in the 75th-percentile county in terms of distance from the line should have approximately 38% lower profits than firms in the 25th-percentile county on average ($-0.102 \times 3.68 = -0.375$). In contrast, average firm profits grew at approximately sixty percent per year during 2004-6. If this was sustained for eighteen years, the cumulative growth in firm profits would be 2,950%. While firm profits did not grow until the later reform era and this crude estimate of cumulative profit growth is likely to be significantly higher than actual firm profit growth over the eighteen year period, it nevertheless illustrates the fact that the implied elasticities between distance from the line and firm profits is relatively small in size.

In panels B and C, we repeat the estimates on samples where the nearest 10% and 20% counties to the line are excluded. The estimates are similar to the main sample estimates. This means that our finding that more firms locate nearer the line is unlikely to reflect a crowding-in effect.

We also examined the effect of distance on the growth of the number of firms, the growth of average firm profits and the 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 for the sake of brevity and because of concerns over the quality of the data for returns to capital. 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 systematically over-state the effect of the line as we move further away from the line. This measurement issue is a generic problem in the Chinese data on firm assets.

Table 7 shows the estimated effects of distance on average household income for agricultural households at the village level. Panel A column (1) shows that distance from the line is negatively correlated with the Gini coefficient for village household incomes. The estimate is statistically significant at the 1% level. This is consistent with the finding that distance from the line is positively correlated with the incomes of the median household, but

negatively correlated with incomes of the poorest and richest households. However, these estimates are not statistically significant.

In Column (5), we estimate the effect of distance on the annual change of the Gini coefficient. It shows that distance from the line is correlated with slower growth in inequality. The estimate is statistically significant at the 1% level. The estimated effects on income growth are statistically indistinguishable from zero.²⁸

In panels B and C, we present the results for restricted samples where we omit 10% and 20% closest counties to the line. The estimates show that the effect on household income inequality is mainly driven by the nearest counties. This might reflect that these areas are the ones that both gain the most in terms of trade opportunities, but also lose most from capital (and, though it is not in our model, human capital) mobility.

6.4 Robustness to Additional Lines

One obvious concern with our strategy regards the relevance of our historical lines. Earlier in this section, we showed that proximity to our lines is positively correlated with proximity to transportation infrastructure such as railroads and coastal routes. However, our estimates also suggested that our lines are uncorrelated with the more recently constructed paved motorways, which have been found by a working paper by Faber (2009) to also affect production and growth. In this section, we test that our main results are robust to controlling for access to such recent transportation infrastructure. Specifically, we directly control for distance to the expanded set of lines constructed by Faber (2009), which he kindly shared with us. This expanded set of lines includes our historical lines of transportation and adds many additional lines to capture recently constructed road networks.

First, we investigate the difference between the historical and expanded lines in terms of how each correlates to transportation networks. Table 3 Panel B shows that on average, the distance from the expanded set of lines is positively correlated to distance from railroads and the distance from the segment city; negatively correlated with the length of highways and

²⁸Arrellano-Bond estimates are presented in Appendix Table A.1 panel B. They are consistent with the main results in showing that distance from the line has little effect on income growth. The estimates are small in magnitude and statistically insignificant.

roads; and uncorrelated with distance from rivers, coastline and country borders. Therefore the key difference between the expanded set of lines and our historical lines is that the distance to the former are negatively correlated with road density (recall that we control for county area), while the distance to the latter are uncorrelated with road density. This is 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 our 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.

In Table 8, we test the robustness our main estimates by running a “horse race” between the historical and expanded sets of lines. Note that the sample size is smaller than the one for our main estimates because Faber’s, (2009) data do not match exactly to ours. Nevertheless, our baseline estimates from using this restricted sample are similar to those from using our full sample. More importantly, the estimates show that our baseline estimates of the effect of historical lines are very robust to controlling for the additional lines. These estimates suggest that the historical lines are indeed the relevant lines to study in our context.

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, a higher number of firms and higher average firm profits. However, these level differences are relatively 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 Fogelian (Fogel, 1992, 1964) interpretation or Huang’s (2008) view that during this period of fast growth, the Chinese government should not have focused so much on building transportation infrastructure. However, they are also consistent

with an alternative explanation where the infrastructure might have brought sizeable benefits for the economy as a whole, but the localization of the gains (and the overall level of the gains) was limited by the lack of factor mobility. The fact that we do not see a strong divergence between well- and poorly connected areas does not rule out the possibility that infrastructure had benefits for all of them, but the lack of factor mobility prevented the gains from being concentrated in relatively better connected areas.

These results should not discourage those who believe that investment in transportation infrastructure can promote economic development. Rather, they highlight the importance of other factors in thinking of the effects of infrastructure on 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, Yuko and Rory Horner**, “World Development Report 2009: Reshaping Economic Geography, by the World Bank,” *Journal of Regional Science*, 2009, 50 (5), 1467–9787.
- Arellano, Manuel and Stephen Bond**, “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,” *Review of Economic Studies*, April 1991, 58 (2), 277–97.
- Atack, Jeremy, Fred Bateman, Michael Haines, and Robert A. Margo**, “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 January 2009.
- Banerjee, Abhijit, Chang-Tai Hsieh, and Nancy Qian**, “The Origins of Private Enterprises in China,” Yale University Working Papers, Yale University 2012.
- , **Esther Duflo, and Nancy Qian**, “The Railroad to Success: The Effect of Access to Transportation Infrastructure on Economic Growth in China,” MIT Working Papers, MIT 2004.
- Cai, Fang**, “Chinese rural reform in 30 years: An analysis in terms of institutional economics,” *Frontiers of Economics in China*, 2010, 5, 150–168. 10.1007/s11459-010-0008-1.
- Donaldson, David**, “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure,” MIT Working Papers, MIT 2010.
- Duflo, Esther**, “The medium run effects of educational expansion: evidence from a large school construction program in Indonesia,” *Journal of Development Economics*, June 2004, 74 (1), 163–197.
- Duranton, Gilles and Diego Puga**, “Micro-foundations of urban agglomeration economies,” in J. V. Henderson and J. F. Thisse, eds., *Handbook of Regional and Ur-*

- ban Economics*, Vol. 4 of *Handbook of Regional and Urban Economics*, Elsevier, April 2004, chapter 48, pp. 2063–2117.
- Eckstein, Alexander**, *China's economic revolution*, Cambridge University Press, 1977.
- Faber, Benjamin**, “Integration and the Periphery: The Unintended Effects of New Highways in a Developing Country,” LSE Working Papers, LSE 2009.
- Feenstra, Robert C.**, “Border Effects and the Gravity Equation: Consistent Methods for Estimation,” *Scottish Journal of Political Economy*, 2002, 49 (5), 491–506.
- Fogel, Robert**, “A Quantitative Approach to the Study of Railroads in American Economic Growth: A Report of Some Preliminary Findings,” *Journal of Economic History*, 1962, 22 (2), 163–197.
- , *Railroads and American Economic Growth: Essays in Econometric History*, Baltimore: Johns Hopkins Press, 1964.
- Glaeser, Edward and Joshua Gottlieb**, “The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States,” *Journal of Economic Literature*, 2009, 47 (4), 983–1028.
- Holtz-Eakin, Douglas, Whitney Newey, and Harvey S Rosen**, “Estimating Vector Autoregressions with Panel Data,” *Econometrica*, November 1988, 56 (6), 1371–95.
- Hsieh, Chang-Tai and Peter J. Klenow**, “Misallocation and Manufacturing TFP in China and India,” *The Quarterly Journal of Economics*, November 2009, 124 (4), 1403–1448.
- Hu, Zuli., Mohsin S. Khan, and International Monetary Fund.**, *Why is China growing so fast? / Zuli Hu and Mohsin S. Khan*, International Monetary Fund, Washington, D.C. :, 1997.
- Huang, Yasheng**, *Capitalism with Chinese Characteristics: Entrepreneurship and the State*, 1 ed., Cambridge University Press, September 2008.

- Jowett, A. John**, “Patterns of literacy in the People’s Republic of China,” *GeoJournal*, 1989, 18, 417–427. 10.1007/BF00772696.
- Keller, Wolfgang and Carol H. Shiue**, “Institutions, Technology, and Trade,” NBER Working Papers 13913, National Bureau of Economic Research, Inc April 2008.
- Martinez-Bravo, Monica, Gerard Padro i Miquel, Nancy Qian, and Yang Yao**, “Accountability under Authoritarianism: The Effect of Electoral Reforms in Rural China,” Yale University Working Papers, Yale University 2010.
- Meng, Xin**, “Introduction: Poverty and labor markets in China,” *Journal of Comparative Economics*, December 2005, 33 (4), 641–643.
- **and Michael P. Kidd**, “Labor Market Reform and the Changing Structure of Wage Determination in China’s State Sector during the 1980s,” *Journal of Comparative Economics*, December 1997, 25 (3), 403–421.
- Michaels, Guy**, “The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System,” *The Review of Economics and Statistics*, 07 2008, 90 (4), 683–701.
- Munshi, Kaivan and Mark Rosenzweig**, “Why is Mobility in India so Low? Social Insurance, Inequality, and Growth,” NBER Working Papers 14850, National Bureau of Economic Research, Inc April 2009.
- Nickell, Stephen J**, “Biases in Dynamic Models with Fixed Effects,” *Econometrica*, November 1981, 49 (6), 1417–26.
- Pong, David**, “Confucian Patriotism and the Destruction of the Woosung Railway, 1877,” *Modern Asian Studies*, 1973, 7 (04), 647–676.
- Spence, D. Jonathan**, *The search for modern China*, Norton, 1990.
- Unger, Jonathan**, *The transformation of rural China Asia and the Pacific*, M.E. Sharpe, 2002.

West, Loraine A. and Yaohui Zhao, *Rural Labor Flows in China*, Berkeley, California:
Berkeley University Press, 2000.

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

Table 1: Descriptive Statistics

Variable	Obs	Mean	Std.Dev.	Variable	Obs	Mean	Std.Dev.
A. Sample 1: County Level GDP (1986-2003)				C. Sample 3: Firms (1993, 2004-06)			
Distance to Historical Line (km)	2744	71.31	64.16	Distance to Historical Line (km)	3663	81.63	87.83
Distance to Railroad (km)	2744	38.81	39.20	Distance to Railroad (km)	3663	49.70	65.26
Length of Highway (km)	2744	6.05	12.77	Length of Highway (km)	3663	5.72	13.80
Length of Paved Roads (km)	2744	83.54	47.16	Length of Paved Roads (km)	3663	92.57	78.33
Distance to River (km)	2744	183.23	137.38	Distance to River (km)	3663	148.43	111.98
Distance to Coastline (km)	2744	424.98	372.41	Distance to Coastline (km)	3663	544.66	405.04
Distance to Country Border (km)	2744	865.14	205.50	Distance to Country Border (km)	3663	705.24	309.89
Distance to Segment City (km)	2744	144.17	81.27	Distance to Segment City (km)	3663	156.13	102.49
County Area (Sqkm)	2744	1918	1134	County Area (Sqkm)	3663	2405	3716
County Population (Individuals)	2744	201347	211227	Number of Firms	3663	81.68	145.83
Per Capita GDP (Constant RMB)	2744	6834	8076	Public Ownership	3663	39.75	79.39
Primary	2266	1968	2424	Mixed Ownership	3663	7.24	16.22
Secondary	2266	2192	3579	Individual Ownership	3663	24.76	80.28
Tertiary	2199	1672	2295	Aggregate (Sum) Profits (Constant 10,000 RMB)	3321	227411	1176004
Per Capita GDP Growth	1879	0.08	0.14	Average Profits	3321	2889	28211
Primary	1644	0.04	0.17	Public Ownership	3233	1997	18066
Secondary	1644	0.10	0.19	Mixed Ownership	1909	12753	88431
Tertiary	1551	0.11	0.17	Individual Ownership	1785	1629	4441
B. Sample 2: Household Income (1987-91, 93, 95-2005)							
Distance to Historical Line (km)	1897	64.79	67.66				
Distance to Railroad (km)	1897	37.82	42.36				
Length of Highway (km)	1897	7.69	14.79				
Length of Paved Roads (km)	1897	89.10	50.94				
Distance to River (km)	1897	129.93	104.50				
Distance to Coastline (km)	1897	475.19	318.46				
Distance to Country Border (km)	1897	739.17	299.00				
Distance to Segment City (km)	1897	133.01	78.90				
County Area (Sqkm)	1897	2058	1261				
Gini	1897	0.28	0.07				
10th Percentile HH Income (Constant RMB)	1897	2729	1322				
50th Percentile HH Income (Constant RMB)	1897	5460	2837				
90th Percentile HH Income (Constant RMB)	1897	11681	11407				
Gini Growth	1782	0.001	0.049				
10th Percentile HH Income Growth	1782	0.018	0.221				
50th Percentile HH Income Growth	1782	0.028	0.162				
90th Percentile HH Income Growth	1782	0.031	0.199				

Notes: Variables are observed at the county and year level. Sample 1 in panel A uses data from *Provincial Statistical Yearbooks*. Sample 2 in panel B uses data from the *National Fixed Point Survey*. Sample 3 in panel C uses data from the *Censuses of Manufacturing Firms*. The geographic data in all samples are computed by the authors using ArcGIS.

Table 2: Distance from the Historical Line and Economic Indicators

Distance to Historical Line by Decile	Per Capita GDP		Annual Per Capita GDP Growth		Median Village Household Income		Gini		Number of Firms		Aggregate Firm Profits	
	Mean (1)	Std. Dev. (2)	Mean (3)	Std. Dev. (4)	Mean (5)	Std. Dev. (6)	Mean (7)	Std. Dev. (8)	Mean (9)	Std. Dev. (10)	Mean (11)	Std. Dev. (12)
0	7035	10613	0.07	0.16	6732	4243	0.30	0.08	111.85	184.24	2451	5202
1	6684	7742	0.08	0.14	5472	1995	0.28	0.06	97.40	171.65	2700	7681
2	7119	9968	0.10	0.12	5010	2428	0.27	0.08	115.95	214.98	1169	6235
3	7079	8365	0.07	0.13	6736	3734	0.29	0.05	96.31	149.72	2571	6170
4	6241	6823	0.08	0.14	4995	1505	0.28	0.06	92.99	135.23	1555	5613
5	7507	8334	0.06	0.14	4475	1819	0.26	0.06	73.43	99.54	2014	4549
6	6958	6789	0.09	0.16	4349	1726	0.26	0.06	93.32	166.48	2639	6407
7	5972	6771	0.08	0.12	5756	3561	0.26	0.07	75.57	101.47	1833	5943
8	5734	6775	0.07	0.13	5383	2429	0.27	0.07	44.37	76.66	9150	85967
9	7994	7505	0.05	0.19	5774	2587	0.28	0.07	16.07	26.42	2800	16706

Notes: Columns (1)-(4) use data from the *Provincial Statistical Yearbooks*. Columns (5)-(8) use data from the *National Fixed Point Survey*. Columns (9)-(12) use data from the *Census of Manufacturing Firms*. Distance is calculated by the authors using ArcGIS.

Table 3: Distance from the Line and Transportation Infrastructure

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent Variables: Distance to Infrastructure						
	Ln Distance to RR	Ln Distance to Segment City	Length of Highway	Length of Paved Roads	Ln Distance to Navigable River	Ln Distance to Coastline	Ln Distance to Country Border
Ln Dist Historical Line	0.206 (0.0615)	0.216 (0.0386)	-0.241 (0.0820)	-0.00812 (0.0405)	-0.0323 (0.0556)	0.161 (0.0632)	-0.0336 (0.0156)
Observations	3127	3127	3127	3127	3127	3127	3127
R-squared	0.085	0.222	0.062	0.339	0.072	0.091	0.054
Ln Distance to Expanded Lines	0.338 (0.0628)	0.343 (0.0408)	-0.460 (0.0747)	-0.0652 (0.0282)	0.0588 (0.0502)	0.0792 (0.0669)	-0.0177 (0.0139)
Observations	2981	2981	2981	2981	2981	2981	2981
R-squared	0.149	0.415	0.163	0.349	0.074	0.076	0.044
Ln Dist Historical Line	0.150 (0.0638)	0.147 (0.0346)	-0.114 (0.0761)	0.00742 (0.0426)	-0.0372 (0.0611)	0.129 (0.0687)	-0.0339 (0.0171)
Ln Distance to Expanded Lines	0.308 (0.0641)	0.313 (0.0378)	-0.437 (0.0782)	-0.0667 (0.0273)	0.0664 (0.0533)	0.0530 (0.0701)	-0.0108 (0.0146)
Observations	2981	2981	2981	2981	2981	2981	2981
R-squared	0.164	0.460	0.170	0.349	0.075	0.088	0.061

Notes: All regressions control for the logarithm of the area of the county, year and province fixed effects. Standard errors are clustered at the county level. These estimates use an unbalanced county-year level panel. Distances to historical lines are computed by the authors. Distance to the expanded set of lines is taken from Farber (2009)

Table 4: The Effect of Distance to the Line on Production Levels

	Dependent Variable: Ln Per Capita GDP							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln Distance to Historical Lines	-0.0617 (0.0286)	-0.0434 (0.0281)	-0.0491 (0.0277)	-0.0474 (0.0279)	-0.0430 (0.0277)	-0.0576 (0.0269)	-0.0697 (0.0274)	-0.0672 (0.0275)
Ln Distance to Segment City		0.0651 (0.232)	-0.0611 (0.266)	-0.0620 (0.264)	-0.0916 (0.276)	-0.0559 (0.276)	-0.184 (0.322)	-0.225 (0.293)
Ln Distance to Segment City ²		-0.0276 (0.0277)	-0.00886 (0.0308)	-0.00727 (0.0309)	-0.00469 (0.0323)	-0.00585 (0.0320)	0.00941 (0.0371)	0.0131 (0.0340)
Ln Distance to Navigable River			0.318 (0.153)	0.305 (0.156)	0.316 (0.150)	0.307 (0.144)	0.362 (0.144)	0.380 (0.141)
Ln Distance to Navigable River ²			-0.0517 (0.0200)	-0.0503 (0.0205)	-0.0503 (0.0198)	-0.0465 (0.0190)	-0.0547 (0.0194)	-0.0549 (0.0188)
Ln Length of Highway				0.00634 (0.146)	-0.0133 (0.140)	-0.0240 (0.131)	-0.0280 (0.135)	-0.0953 (0.131)
Ln Length of Highway ²				0.00329 (0.0410)	0.0111 (0.0393)	0.0123 (0.0374)	0.0121 (0.0384)	0.0298 (0.0371)
Ln Length of Road					0.121 (0.109)	0.0921 (0.112)	0.116 (0.104)	0.0973 (0.103)
Ln Length of Road ²					-0.0252 (0.0174)	-0.00799 (0.0171)	-0.0113 (0.0163)	-0.00789 (0.0160)
Ln Area						-1.583 (0.699)	-1.464 (0.663)	-1.436 (0.666)
Ln Area ²						0.0973 (0.0504)	0.0910 (0.0483)	0.0867 (0.0484)
Ln Distance to Coastline							-0.216 (0.223)	-0.178 (0.214)
Ln Distance to Coastline ²							0.0132 (0.0263)	0.0104 (0.0254)
Ln Distance to Country Border								-16.99 (6.201)
Ln Distance to Country Border ²								1.279 (0.468)
Observations	2744	2744	2744	2744	2744	2744	2744	2744
R-squared	0.818	0.826	0.833	0.833	0.835	0.847	0.850	0.853

Notes: All regressions control for year and province fixed effects. Standard errors are clustered at the county level. These estimates use an unbalanced county-year level panel. GDP data are from *Provincial Statistical Yearbooks*. All geographic variables are computed by the authors.

Table 5: The Effect of Distance to the Line on Production Levels and Growth

	Dependent Variables: Per Capita GDP							
	Ln(Per Capita GDP)				Annual Growth in Ln(Per Capita GDP)			
	All (1)	Primary (2)	Secondary (3)	Tertiary (4)	All (5)	Primary (6)	Secondary (7)	Tertiary (8)
Ln Dist Historical Line	-0.0672 (0.0275)	-0.0295 (0.0213)	-0.0969 (0.0464)	-0.0784 (0.0329)	-0.00207 (0.00343)	-0.000325 (0.00538)	-0.00719 (0.00524)	-0.00123 (0.00380)
Observations	2744	2266	2266	2199	1110	1018	1018	899
R-squared	0.853	0.871	0.723	0.843	0.206	0.226	0.183	0.183
Ln Dist Historical Line	-0.0578 (0.0381)	-0.0477 (0.0396)	-0.0513 (0.0617)	-0.0458 (0.0413)	-0.000664 (0.00508)	0.00969 (0.00711)	-0.00769 (0.00849)	-0.00318 (0.00727)
Observations	2498	2045	2045	1983	1009	919	919	809
R-squared	0.856	0.871	0.724	0.847	0.229	0.286	0.172	0.178
Ln Dist Historical Line	-0.0886 (0.0511)	-0.100 (0.0495)	-0.101 (0.0829)	-0.0632 (0.0536)	-0.000394 (0.00543)	0.00930 (0.0113)	-0.00285 (0.00986)	-0.00396 (0.00627)
Observations	2195	1784	1784	1723	882	798	798	692
R-squared	0.864	0.881	0.726	0.855	0.257	0.306	0.163	0.177

Notes: All regressions control for the full set of baseline controls in Table 4 column (8). The growth regressions in columns (5)-(8) also control for the one and two year lags of sector-specific per capita GDP levels. Standard errors are clustered at the county level. The sample is an unbalanced county-year level panel. The GDP data are from *Provincial Statistical Yearbooks*. All geographic variables are computed by the authors.

Table 6: The Effect of Distance to the Line on Firm Location and Profits

	Dependent Variables							
	Ln Number of Firms				Average Firm Profits			
	(1) All	(2) Public	(3) Mixed	(4) Individual	(5) All	(6) Public	(7) Mixed	(8) Individual
Ln Distance to Historical Lines	-0.084 (0.021)	-0.053 (0.025)	-0.082 (0.033)	-0.117 (0.032)	-0.102 (0.040)	-0.090 (0.047)	-0.074 (0.068)	-0.022 (0.044)
Observations	3321	3321	3321	3321	2763	2416	1642	1503
R-squared	0.650	0.685	0.706	0.787	0.450	0.324	0.206	0.350
Ln Distance to Historical Lines	-0.118 (0.033)	-0.085 (0.035)	-0.093 (0.053)	-0.139 (0.053)	-0.147 (0.066)	-0.157 (0.076)	-0.097 (0.095)	-0.051 (0.072)
Observations	2994	2994	2994	2994	2477	2151	1460	1343
R-squared	0.658	0.699	0.697	0.779	0.444	0.327	0.214	0.357
Ln Distance to Historical Lines	-0.145 (0.045)	-0.067 (0.045)	-0.133 (0.067)	-0.153 (0.071)	-0.172 (0.085)	-0.055 (0.090)	-0.012 (0.122)	-0.021 (0.099)
Observations	2663	2663	2663	2663	2196	1897	1288	1182
R-squared	0.666	0.700	0.705	0.774	0.451	0.336	0.226	0.348

Notes: All regressions control for the full set of baseline controls in Table 4 column (8). Standard errors are clustered at the county level. The sample is an unbalanced county-year level panel. The data for the dependent variables are from the *Censuses of Manufacturing Firms*. All geographic variables are computed by the authors.

Table 7: The Effect of Distance to the Line on Income Inequality

	Dependent Variables: Household Income Distribution							
	Ln(HH Income) by Percentile				Annual Growth Rate			
	(1) Gini	(2) 10th	(3) 50th	(4) 90th	(5) Gini	(6) 10th	(7) 50th	(8) 90th
Ln Distance to Historical Lines	-0.00698 (0.00338)	-0.00988 (0.0277)	0.00421 (0.0242)	-0.00179 (0.0286)	-0.00182 (0.000626)	0.00136 (0.00336)	-0.000876 (0.00212)	-0.00111 (0.00278)
Observations	1897	1897	1897	1897	1533	1533	1533	1533
R-squared	0.277	0.466	0.605	0.591	0.036	0.124	0.171	0.096
Ln Distance to Historical Lines	-0.00383 (0.00530)	0.0263 (0.0391)	0.0392 (0.0325)	0.0337 (0.0418)	-0.00172 (0.000908)	0.00138 (0.00477)	0.00246 (0.00364)	-0.00175 (0.00484)
Observations	1722	1722	1722	1722	1392	1392	1392	1392
R-squared	0.260	0.493	0.610	0.591	0.042	0.145	0.200	0.099
Ln Distance to Historical Lines	-0.000974 (0.00765)	0.0689 (0.0507)	0.0854 (0.0419)	0.0853 (0.0557)	-0.000410 (0.00104)	0.00459 (0.00602)	0.00640 (0.00485)	0.00731 (0.00514)
Observations	1531	1531	1531	1531	1238	1238	1238	1238
R-squared	0.268	0.507	0.627	0.615	0.044	0.135	0.185	0.089

Notes: All regressions control for the full set of baseline controls in Table 4 column (8). The regression in column (5) also controls for the 1 and 2 year lag of the gini, and the regressions in columns (6)-(8) control for the 1 and 2 year lags of the relevant income levels. Standard errors are clustered at the county level. The sample is an unbalanced panel of villages. Income data are from the *National Fixed Point Survey*. All geographic data are computed by the authors.

Table 8: Historical Lines and More Lines

	Dependent Variables			
	(1) Ln PC GDP	(2) Ln PC GDP Primary	(3) Ln PC GDP Secondary	(4) Ln PC GDP Tertiary
A. Baseline				
Ln Distance to Hist Line	-0.0788 (0.0279)	-0.0278 (0.0225)	-0.120 (0.0464)	-0.0873 (0.0335)
Observations	2605	2142	2142	2096
R-squared	0.862	0.862	0.745	0.849
B. Control for Expanded Lines				
Ln Distance to Hist Line	-0.0792 (0.0282)	-0.0277 (0.0225)	-0.121 (0.0471)	-0.0880 (0.0341)
Ln Distance to Expanded	-0.0224 (0.0324)	0.0112 (0.0294)	-0.0495 (0.0479)	-0.0433 (0.0424)
Observations	2605	2142	2142	2096
R-squared	0.862	0.862	0.746	0.850
C. Baseline				
	Gini	Ln Inc 10th	Ln Inc 50th	Ln Inc 90th
Ln Distance to Hist Line	-0.00620 (0.00364)	0.0337 (0.0286)	0.0396 (0.0267)	0.0370 (0.0322)
Observations	1070	1070	1070	1070
R-squared	0.329	0.560	0.612	0.568
D. Control for Expanded Lines				
Ln Distance to Hist Line	-0.00712 (0.00402)	0.0390 (0.0328)	0.0433 (0.0304)	0.0428 (0.0352)
Ln Distance to Expanded	-0.00429 (0.00599)	0.0248 (0.0456)	0.0173 (0.0405)	0.0273 (0.0444)
Observations	1070	1070	1070	1070
R-squared	0.331	0.561	0.613	0.569
E. Baseline				
	Ln Total # Firms	Ln # Public Firms	Ln # Mixed Firms	Ln # Individual Firms
Ln Distance to Hist Line	-0.094 (0.025)	-0.064 (0.030)	-0.074 (0.039)	-0.141 (0.037)
Observations	2704	2704	2704	2704
R-squared	0.677	0.644	0.702	0.776
F. Control for Expanded Lines				
Ln Distance to Hist Line	-0.094 (0.025)	-0.064 (0.030)	-0.074 (0.039)	-0.142 (0.037)
Ln Distance to Expanded	-0.013 (0.026)	-0.002 (0.035)	0.007 (0.040)	0.037 (0.045)
Observations	2704	2704	2704	2704
R-squared	0.677	0.644	0.702	0.776

Notes: All regressions control for the full set of baseline controls in Table 4 column (8). Standard errors are clustered at the county level. The sample is an unbalanced county-year level panel. The GDP data used in panels A-B are from *Provincial Statistical Yearbooks*. The income data used in Panels C-D are from the *National Fixed Point Surveys*. The firm data used in panels E-F are from the *Censuses of Manufacturing Firms*. The distance to the historical line is calculated by the authors. The distance to the expanded set of lines is taken from Farber (2009).

Figure 1: Lines and Transportation Infrastructure

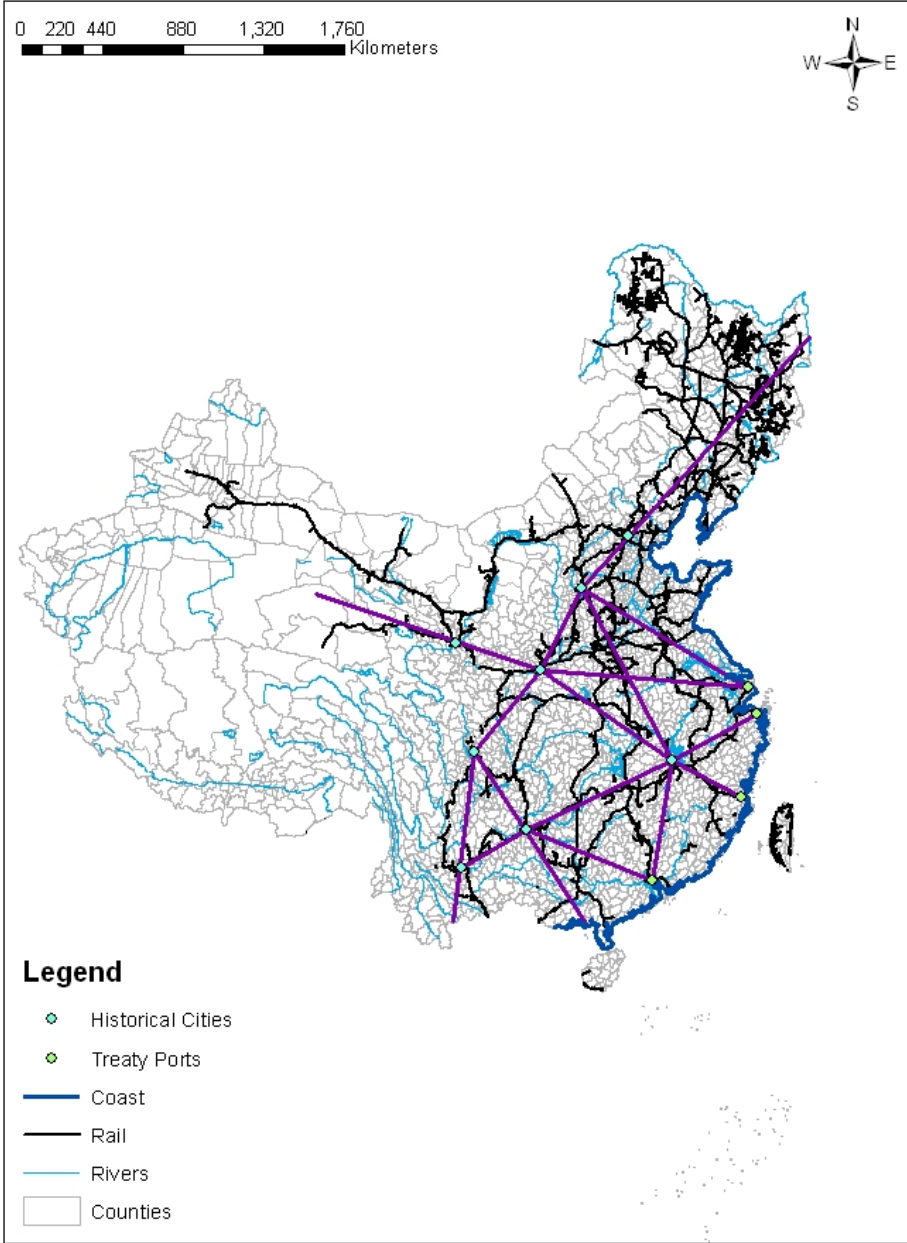


Figure 2: Counties with GDP data from the *Provincial Statistical Yearbooks*

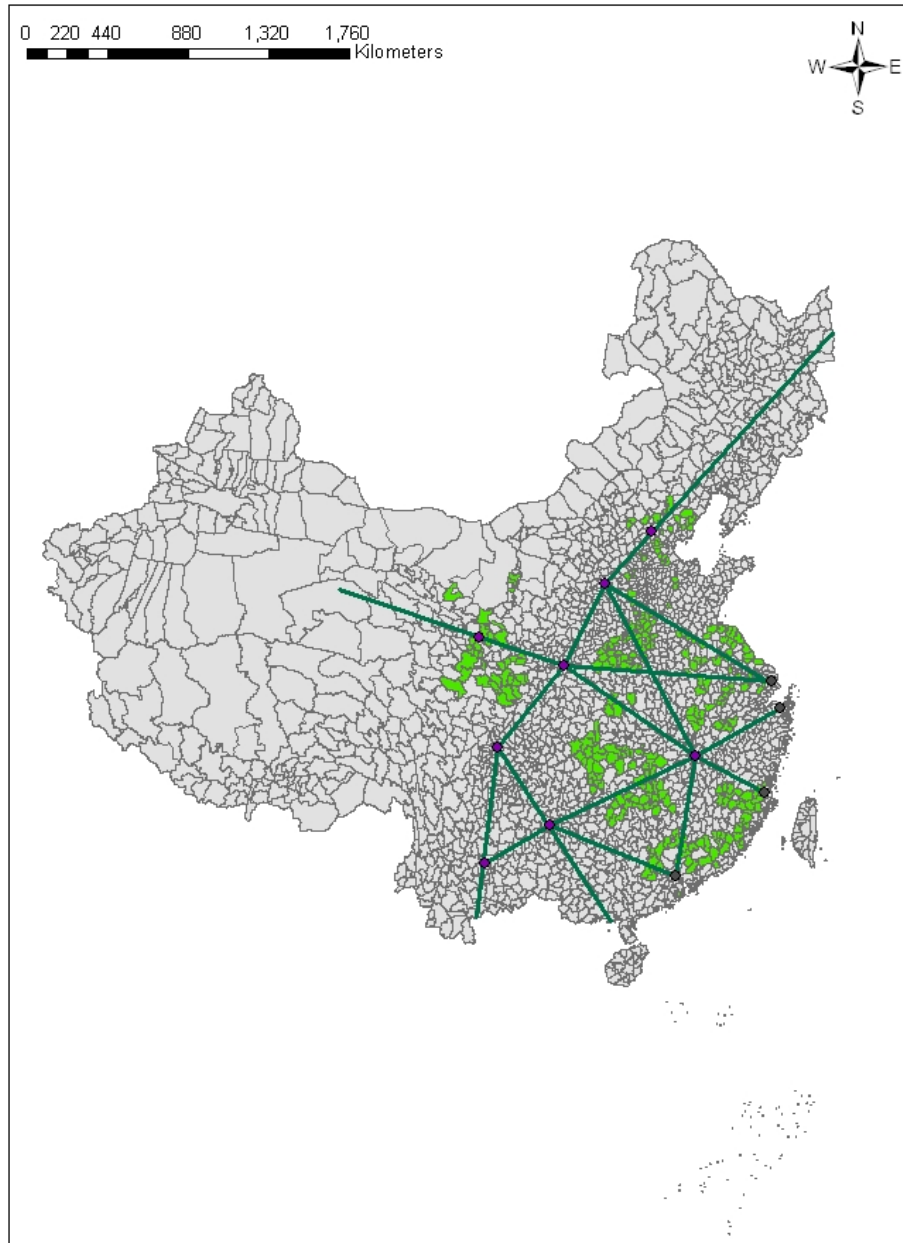


Figure 3: Counties with Firm data from the *Censuses for Manufacturing Firms*

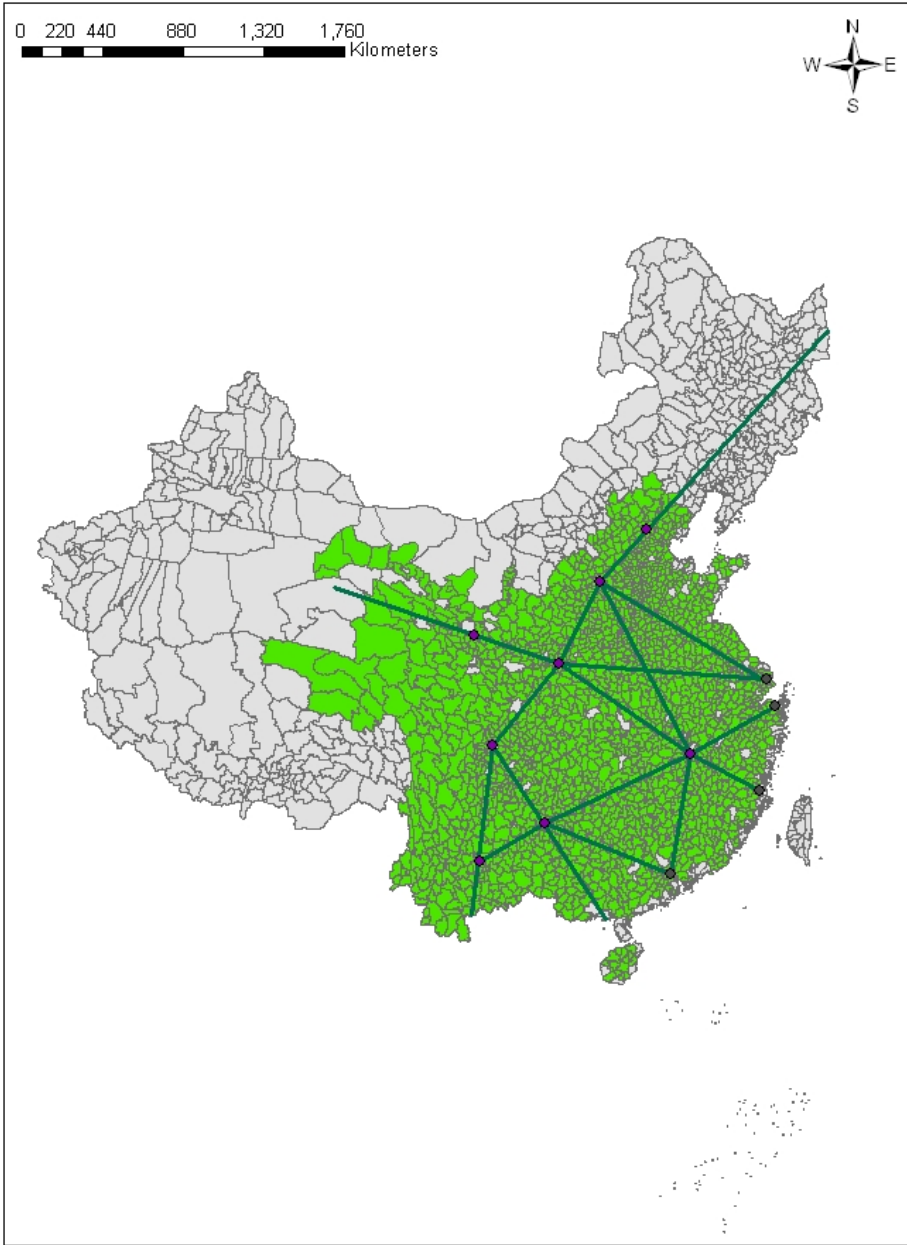


Figure 4: Counties with Income data from the *National Fixed Point Survey*

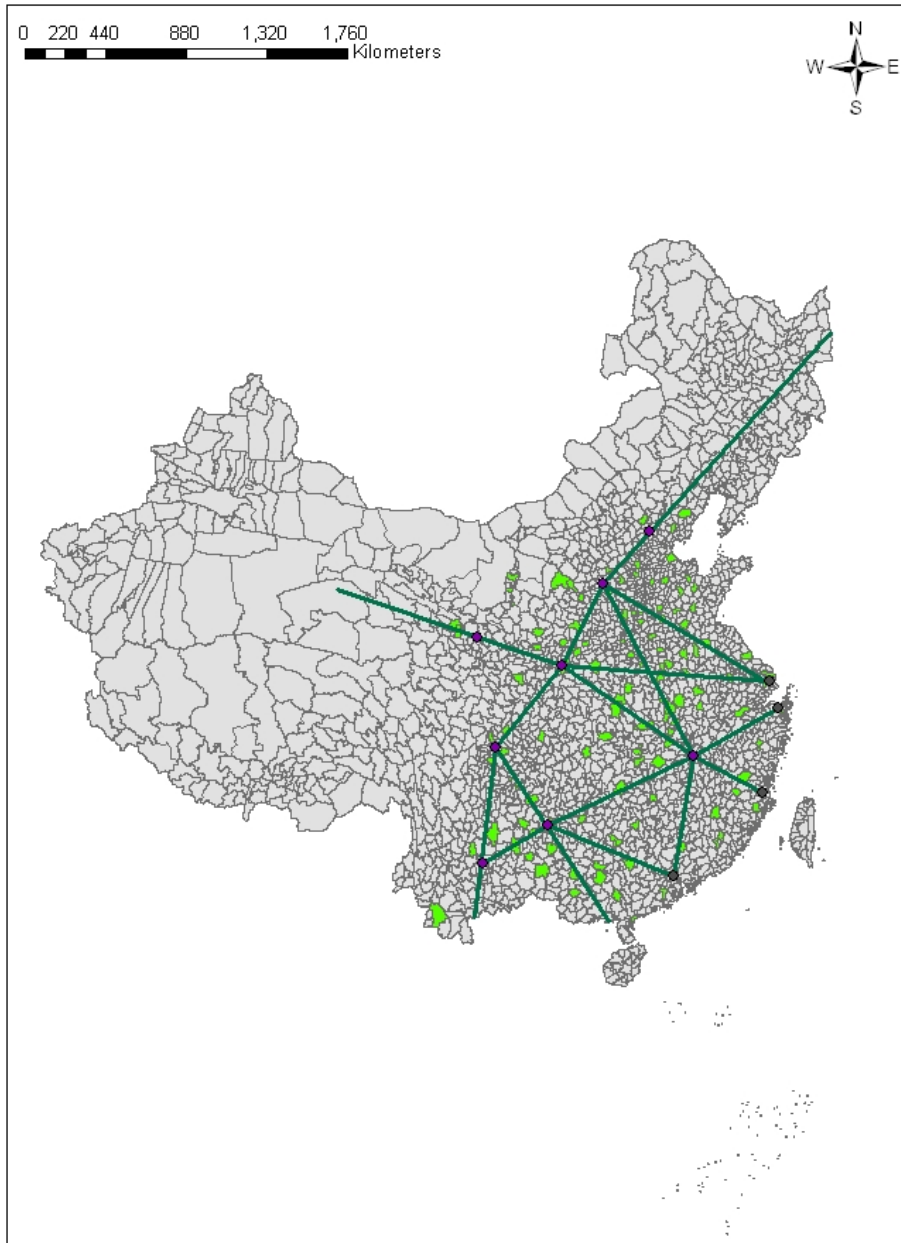


Table A.1: The Effect of Distance to the Line on Production – Dynamic Panel Estimation

		Dependent Variables: Growth											
All Sectors		Primary			Secondary			Tertiary					
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)		
1 Lag	2 Lags	3 Lags	1 Lag	2 Lags	3 Lags	1 Lag	2 Lags	3 Lags	1 Lag	2 Lags	3 Lags		
A. Per Capital GDP Growth													
Ln Distance to Hist Lines	0.00474 (0.00325)	0.00484 (0.00328)	-0.000276 (0.00411)	-0.000768 (0.00415)	-0.000151 (0.00415)	0.00678 (0.00497)	0.00693 (0.00501)	0.00695 (0.00503)	0.00280 (0.00330)	0.00279 (0.00330)	0.00280 (0.00330)		
Observations	1373	1373	1281	1281	1281	1281	1281	1281	1175	1175	1175		
Number of Counties	258	258	258	258	258	258	258	258	245	245	245		
B. Rural Household Income Growth													
10th													
Ln Distance to Hist Lines	0.00215 (0.00340)	0.00216 (0.00346)	0.000076 (0.00211)	-0.0000212 (0.00214)	-0.0000212 (0.00214)	-0.00127 (0.00297)	-0.00128 (0.00302)	-0.00128 (0.00302)	-0.00128 (0.00302)	-0.00128 (0.00302)	-0.00128 (0.00302)		
Observations	1655	1655	1655	1655	1655	1655	1655	1655	1655	1655	1655		
Number of Counties	122	122	122	122	122	122	122	122	122	122	122		

Note: All regressions control for the full set of baseline controls in Table 4 column (8). In panel A, the lagged dependent variable is instrumented using 1, 2 or 3 lags of the per capita GDP level for the relevant sector. In panel B, the lagged dependent variable is instrumented with 1, 2, or 3 lags of the relevant income level. The standard errors are clustered at the county level. The sample is an unbalanced panel of counties. The data in panel A are from *Provincial Statistical Yearbooks*. The data in panel B are from the *National Fixed Point Survey*. All geographic data are calculated by the authors.