Highways, Market Access, and Urban Growth in China *

Nathaniel Baum-Snow, University of Toronto
Vernon Henderson, London School of Economics
Matthew A. Turner, Brown University   Qinghua Zhang, Peking University
Loren Brandt, University of Toronto

February 28, 2017

Abstract

We investigate the effects of the recently constructed Chinese national highway system on local economic outcomes. On average, roads that improve access to local markets have small or negative effects on prefecture economic activity and population. However, these averages mask a distinct pattern of winners and losers. With better regional highways, economic output and population increase in regional primates at the expense of hinterland prefectures. Highways also affect patterns of specialization. With better regional highways, regional primates specialize more in manufacturing and services, while peripheral areas lose manufacturing but gain in agriculture. Better access to international ports promotes greater population, GDP, and private sector wages on average, effects that are probably larger in hinterland than primate prefectures. An important policy implication is that investing in transport infrastructure to promote growth of hinterland prefectures has the opposite effect, causing them to specialize more in agriculture and lose economic activity.

*We are grateful to International Growth Centre for generously funding this research. We received helpful comments from Gerald Carlino, Edward Glaeser, Samuel Marden, Daniel Sturm and Junfu Zhang and many seminar participants. We thank Ying Chen for research assistance.

1 Introduction

Between the late 1990s and 2010, China constructed an extensive modern highway network. We investigate the effects of this highway network on China's economic geography. In particular, we examine the effects of the highway network on the spatial distribution of population, GDP, GDP per person, wages, the composition of output and international trade flows, all around 2010.

Our investigation is organized around three main questions. First, what are the effects on a prefecture of better access to regional domestic markets? Second, what are the effects on a prefecture of better access to international markets? Third, how do the effects of market access depend on where a prefecture is in the urban hierarchy?

Our investigation faces two main problems. First, answering these questions requires making causal statements about the effects of the highway network. To estimate casual effects, we rely on plausibly quasi-random variation from the 1962 road network, a network that predates China's transformation into a market-oriented economy. Second, we must measure both a prefecture's location in the urban hierarchy and how the highway system affects 'access' to regional and international markets.

To measure a prefecture's location in the urban hierarchy we define 'regional primate prefectures' as the largest prefectures within about a one day drive and 'hinterland prefectures' otherwise. The scale of this definition, 'about a one day drive', is determined empirically using a technique like those used to test for structural breaks in time series data.

Measuring market access is more difficult. Theoretically founded definitions of market access are fundamentally recursive. If improved access to prefecture B by prefecture A increases the size of prefecture A's economy, then the converse relationship should also be true. However, this implies that shocks to prefecture A's economy effect prefecture A recursively through prefecture B. This raises obvious problems for the estimation of the causal effects of market access. We investigate three solutions to this problem.

First, we skirt the problem by considering measures of access that depend only on the highway network. Specifically, to measure access to the regional domestic economy, we measure the quantity of highways within 450km of each prefecture; and to measure access to international markets we calculate travel time to a major international port along the highway network. Since these measures do not depend on nearby economic activity, they avoid the recursion problem.

Our second measure of market access is the inverse travel time weighted sum of economic activity around a prefecture, which we call 'market potential'. Unlike our quantity based measure, this market potential measure allows connections to larger markets to be more important than connections to smaller markets. This intuitive property, however, comes at the price of introducing the recursion problem described above.

Our third measure of market access derives from a Ricardian model of the sort now common in the economic geography literature (e.g., Donaldson and Hornbeck (2016) or

Tombe and Zhu(2015)), and an almost identical measure can be derived from an NEG model. This measure allows a more nuanced description of market access than market potential, has explicit microeconomic foundations and allows a unified treatment of domestic and international access. However, like our market potential measure, our structurally founded market access measure is subject to the recursion problem.

To estimate causal effects of these measures of access, we must address the possibility that regional roads are assigned to prefectures on the basis of unobserved determinants of economic activity. This is a conventional endogeneity problem, and for our highway network measures we address it by relying on quasi-random variation in the 1962 road network. For our market potential or access measures, we resolve the recursion problem by relying on the same quantity based instruments we use for our quantity based measures or travel time measures of roads. Because the instruments do not involve measures of economic activity, their use resolves the structural endogeneity problem that arises from the recursive nature of the market potential and market access variables.

Our investigation leads to three main findings. First, improved access to domestic markets reduces prefecture population, population growth and wages paid by private sector firms. These findings are robust to econometric technique and to the definition of market access. Improved access to domestic markets also reduces prefecture GDP, but this conclusion is less robust to variations in technique. On average, prefectures with better access to domestic markets lose population and probably economic activity. That is, economic activity disperses as there are more roads around a prefecture.

This average effect masks differences in how roads affect prefectures at different positions in the regional hierarchy. Better access to local markets causes increases to regional primates' populations, GDP and wages. On the other hand, hinterland prefectures shrink with better access to domestic markets, and they become relatively more specialized in agriculture at the expense of manufacturing and services. These findings are robust to econometric technique. In addition, the effects of improved market access on regional primates truly seem to reflect a prefecture's position in the urban hierarchy. They do not reflect a prefecture's rank in the national size distribution or whether the prefecture is a nodal point on the highway network or a provincial capital.

Finally, we find that for all prefectures, better access to international ports increases population, population growth between 1990 and 2010, GDP and private sector wages. These effects are less important for primate than hinterland prefectures. The effect of access to international markets operates equally on both the value of imports and the value of exports. International trade decreases with access to regional domestic markets and this effect is the same in primate and hinterland prefectures. This suggests that access to regional domestic markets leads to substitution of domestic for foreign goods and inputs.

Our findings suggest that the highway system has a profound and complicated effect on the economic geography of China. Overall marginal effects involve clear reshuffling of economic activity to concentrate people in regional primates and prefectures well connected to international ports at the expense of other prefectures. With this migration comes an increase in output in the regional primates, manufacturing in particular, while hinterland prefectures shrink and specialize in agriculture. The effects on productivity measures such as wages and GDP per capita are generally less sharp. Results suggest improved access to international markets increases a prefecture's productivity, but perhaps not so for domestic access.

These results are important for several reasons. First, billions of dollars of developing world transportation infrastructure is under construction or consideration. About 20% of World Bank lending supports transport infrastructure projects, more than for poverty reduction. Moreover, with almost half of the population of developing countries now living in cities, and this share rising rapidly, a better understanding of the role transportation infrastructure plays in urban growth is central for informing development policy.

Understanding the effects of improved connections between hinterland cities and regional or other centres is particularly important in China. The 2005 Reform and Development Commission focused on the development of the road network well beyond 2010, with investments under titles such as 'Developing the West' or 'Revitalizing the Northeast,' while the 12th and 13th 5-year plans emphasize the development of poor hinterland regions through a vast expansion in road connections. Our results suggest that these policies may not help these areas retain population, but instead may accelerate their decline. While this may not result in welfare losses given migration responses, it is the opposite of what is intended.

Second, to our knowledge, ours is the first attempt to distinguish the effects of an expressway system on access to local versus international markets. In China, at least, the effects of local and international access are different, and are different for different types of cities. In addition, for the Chinese highway network, the effect of access to regional markets is more important than the effects of international access in determining winners and losers.

Third, to our knowledge, we are the first to provide econometric evidence for an 'urban hierarchy' at the regional level. This finding has several important implications for the study of economic geography in general and transportation infrastructure in particular.

In a seminal paper and subsequent generalization, Krugman (1991) and Fujita, Krugman and Venables (1999) develop a model of economic geography where a decline in trade costs may lead to a decline in hinterland population because the reduction in trade costs allows production to migrate to the core. While the existing empirical literature hints at the presence of urban hierarchies,[1] we provide econometric evidence for an 'urban hierarchy' at the regional level, and hence provide evidence in support of this important class of models. Our findings are also relevant for the theoretical literature describing central place theory. Central place theory originates with Christaller (1933) and consists primarily of the conjecture that in any given region there should be a dominant city, the 'central

[1] For example, Ghani, Goswami and Kerr (2016) find that India's new highway network favored nodal cities while Redding and Sturm (2008) find that small German cities were more adversely affected by German reunification than large ones.

place', that produces a full range of goods for sale to smaller more specialized cities, which may produce goods for still smaller cities in turn. This conjecture forms the basis for a theoretical literature that attempts to rationalize this geography from formal foundations. As noted above, Krugman (1991) provides such foundations in a geography consisting of two discrete locations, while Fujita, Krugman and Mori (1999) and Tabuchi and Thisse (2011) develop specific general equilibrium models of such urban hierarchies along a line and around a circle. Our finding that reductions in transportation costs favor regional primates provides an important comparative static that can be used as a test of competing models of economic geography.

Our work also relates to both the empirical and theoretical literatures on effects of national transport systems, e.g.; Donaldson (2015), Donaldson and Hornbeck (2016), Alder (2016), Sotelo (2015), Allen and Arkolakis (2014), Bartelme (2015) and Fajgelbaum and Redding (2014). After experimenting with calibrations of standard versions of these models, we conclude that our results pose a challenge for existing structural models. In particular, the bifurcation of primate and hinterland prefecture outcomes that follow from the Chinese road network is difficult to replicate with standard models where returns to scale are small, where locations are only differentiated by productivity and amenity shifters and where the land share in production is common across sectors. Fundamentally, the presence of an urban hierarchy appears to require either returns to scale that are important enough to permit multiple equilibria, as in Krugman (1991), or else an important role for prefecture abundance of land or natural resources. The recent literature generally assumes returns to scale are 'small enough' to rule out multiple equilibrium, and factor abundance plays at most a small role in most structural models applied to transportation.

Because our main results are difficult to rationalize with standard models, we do not present structural estimates or explicit welfare calculations. Instead, we focus on establishing facts about the impacts of the expressway system on China's economic geography. Our object is to identify what forces determine the winners and losers from the reshuffling of economic activity caused by the expressway system, as well as to provide facts relevant to future development of structural models.

Our findings also allow us to reconcile the apparent contradictions in the literature investigating the effect of roads and highways in China. Faber (2014) concludes that rural Chinese prefectures are *harmed* by proximity to new highways. Banerjee, Duflo and Qian (2012) conclude that proximity to a highway or railroad is *beneficial* for an average Chinese county. Baum-Snow, Brandt, Henderson, Turner and Zhang (2016) conclude that within prefecture highways have no overall effect on the level of prefecture population or economic activity. Our results allow us to explain these results as a consequence of sampling. Faber (2014) deliberately oversamples rural prefectures, Baum-Snow, Brandt, Henderson, Turner and Zhang (2016) consider the universe, while Banerjee, Duflo and Qian (2012) appear to oversample primate prefectures. Thus, these seemingly contradictory results can be explained if roads cause a shift from hinterland to primate prefectures, but have small effect on average. This is exactly what we find. In addition, contrary to counterfactual predictions in Tombe and Zhu (2015), we find that highways do not centralize economic activity, but move it from hinterland prefectures to regional primates. Also contrary to these counterfactual predictions, the data suggest that access to international markets is more important for growth of the average prefecture than is access to domestic markets.

Finally, we contribute to the literature on the role of transportation infrastructure and trade costs in economic development. Topalova and Khandelwal (2011) provide evidence that lower trade costs has fostered innovation through competition in India. Innovative ideas (Alvarez, Buera and Lucas, 2013; Buera and Oberfeld, 2014) are additional mechanisms through which trade may promote growth. Storeygard (2016) and Bird and Straub (2015) provide related reduced form evaluations of the effects of road networks on cities in Africa and Brazil respectively.

2 Context and Data

The Chinese context is especially well-suited for our investigation. First, China is large and geographically varied enough to permit the emergence a large number of regional primate cities and their hinterlands. Second, the policy intervention is enormous. China had essentially no limited access highways in 1990 and Chinese prefectures experienced large variation in the expansion of the local network since 1990. In 1990, intercity roads had at most two lanes with unrestricted access and, in many places, were not even paved. Almost all goods moved by rail or river, with less than 5% of freight ton-miles moved by road. By 2010 China had constructed an extensive intercity highway network, including the national expressway system. Construction started slowly, with only a few highways complete by 2000, but sped up so that a network serving the whole nation and carrying over 30% of freight ton-miles was in place before 2010, the year for which we generate most results. This highway construction program has resulted in considerable variation across prefectures in how well connected they are to nearby hinterland markets and coastal ports.

In addition, the unique Chinese historical context allows us to construct plausibly exogenous instruments for transport networks on the basis of an historical road network from 1962. In 1962, roads existed primarily to move agricultural goods to local markets within prefectures while railroads existed to ship raw materials and manufactures between larger cities and to provincial capitals according to the dictates of national and provincial annual and 5-year plans. Lyons (1985, p. 312) states: 'At least through the 1960s most roads in China (except perhaps those of military importance) were simple dirt roads built at the direction of county and commune authorities. According to Chinese reports of the early 1960s, most such roads were not fit for motor traffic and half of the entire network was impassable on rainy days.' Lyons also notes that average truck speeds were below 30 km/hr due to poor road quality. However for our purposes, historical roads provide rights-of-way facilitating lower cost highway construction over or alongside old roads, all of which has taken place since 1990.

Figures 1a and 1b show the national road networks in 1962 and 2010. We use the 1962 network to construct instruments for 2010 roads and travel costs. These travel costs assume speeds of 25 kph on local highways and 90 kph on expressways. The lightly shaded region is our study area. We use the 285 prefectures in this area as our primary estimation sample.[2]

2.1 Population and Internal Migration

Because prefecture population is one of our outcome variables, it is important to understand the history of interregional population mobility in China. Before 2000, with the exception of a few coastal cities, cities hosted few inter-province or even inter-prefecture migrants. Migration was limited by the hukou system, which regulated and restricted migration between prefectures and imposed penalties for un-licensed migration. These restrictions were lifted in stages starting in the late 1990s and by the early 2000s unlicensed migration was no longer illegal, although the hukou system continues to restrict migrants' access to formal housing markets, schools, health care, and social security (Chan, 2005), particularly in larger cities. Despite restrictions, a great deal of migration has occurred since 1990. In 1990 China's population was about 30% urban, a share that rose to about 50% by 2010.

Chinese administrative geography dictates the spatial units that we use in our analysis. Provinces are broken into prefectures and prefectures into counties. Our analysis considers 285 prefectures in Han China, about half the land area of China. We omit minority areas for data and contextual reasons and one island prefecture. Our study area contains almost 90% of China's population. Over the course of our study period, the boundaries of a number of counties and prefectures changed, requiring painstaking work to establish county level correspondences over time and time-consistent spatial units. We index all data to prefectures defined as of 2010.

2.2 Outcomes and Controls

We are interested in understanding how highways influence the spatial distribution of economic activity. Because economic geography typically predicts the effects of trade costs on population, output and wages, they are our primary outcomes of interest. Specifically, log 2010 population and log 2010 GDP are our primary outcomes, with log 2010 GDP per capita and log 2007 private sector wages as alternative measures of output per worker. As a robustness check, we consider 1990-2010 population growth rates. Data quality precludes an examination of wage and GDP measures from earlier periods, and hence of changes in those outcomes. To investigate the mechanisms through which roads affect economic activity and population, we also look at effects on industrial composition in 2008-2010 and international trade flows in 2007.

[2] For the purpose measuring infrastructure, we include roads within China that are outside our study area.

We use data from the 1982, 1990, and 2010 population censuses to calculate prefecture population and employment by sector plus various demographic control variables. The 1990 and 2010 data are 100% counts aggregated to rural counties, county cities and prefecture cities or urban districts. The 1982 data are our own aggregation of microdata drawn from a 1% sample for the same geographies. 2010 is from the University of Michigan's Online China Data Archive, which covers prefectures, prefecture cities and rural counties. To calculate industrial composition, we use disaggregated employment data from the 2008 Economic Census. Wage data comes from the 2007 Survey (actually a census) of Medium and Large Industrial Firms and are calculated as total compensation per worker by establishment. We also use data on international trade flows to and from each prefecture derived from customs records.

Figures 1c and 1d show heat maps of 2010 and population respectively, in which lighter shades indicate higher ranks. These figures show that the more central areas of the country have greater population and are more prosperous, with the more peripheral regions less so. One central goal of our analysis is to evaluate the extent to which road infrastructure has contributed to these spatial patterns of economic activity.

2.3 Regional Primate Cities

To investigate the role of the urban hierarchy, we must first develop a statistical description of it. We base our description of the urban hierarchy around the idea of 'regional primates' and their associated 'hinterlands'. We define a prefecture to be a regional primate if, on the basis of 1982 population and travel time over the 1962 road network at 90 kph, it has the largest population within a 360 minute drive. We choose to measure population and the road network as of 1982 and 1962 respectively in order to avoid the possibility that regional primacy responds to highway treatments. We choose the 360 minute scale on the basis of a 'structural break test' that we discuss below, but note that this seems like an intuitively reasonable time as it amounts to about one day's drive. Regional primates are outlined in black in Figure 1e. They are spread throughout the country, but cluster in areas with low road density. Regional primates have larger population on average than other locations but small prefectures are well-represented. Indeed, 27% of primate prefectures are below the median population of primate prefectures of 2.8 million. Unsurprisingly, the top four 1982 population prefectures are all primates.

As a robustness check we also consider a related continuous measure of regional primacy. To calculate this measure, we first identify regional primates as above. Given this classification, for each prefecture we calculate the ratio of its 1982 population to the 1982 population of its regional primate. Thus, all regional primates are ranked one, and hinterland cities receive values strictly between zero and one. This measure refines the regional primate indicator by preserving more information about the size of each prefecture relative to its neighbors. We obtain qualitatively similar results for each of the two measures of regional primacy.

2.4 Roads and Travel Time

To describe the Chinese road and railroad network, we digitize a series of large scale national paper maps. Using the resulting digital maps, we calculate travel times between each pair of prefecture cities over the highway network in each year. To understand the potential importance of links to the international economy, we also calculate travel times over the road network from each prefecture city to the nearest of the nine most important international ports.[3] We rely most heavily on the 1962 and 2010 maps seen in Figures 1a and 1b.

The paper maps on which our digital maps are based were printed by the same publisher, drawn using the same projection and have similar legends. To the extent possible, our data describe consistent sets of roads over time. However, the growth and improvement of China's road network was so dramatic that roads that were important enough to merit inclusion on the 1990 map probably bear little resemblance to roads that meet this standard in 2010, even if both roads receive the same designation in the legend. Thus, we are reluctant to exploit the time series variation in our measures of highways. It is this data limitation together with incomplete GDP information for 1990 that motivate our focus on cross-sectional research designs. With this said, our data do permit an examination of population changes from 1990 to 2010, and we present these results.

The 2010 map describes limited access highways and two classes of smaller roads, on which we assume travel speeds of 90 kph and 25 kph respectively. This allows us to calculate pairwise travel times between any pair of prefecture cities and between each prefecture city and the nearest of the nine international ports described above.[4]

Our measures of market access and market potential, defined below, depend on iceberg trade costs calculated from these pairwise travel times. That is, to deliver one unit of any variety in i from j we must ship $\tau_{ij} \geq 1$ units of that variety. To calculate τ_{ij}, we use

$$\tau_{ij} = 1 + 0.004\rho(\text{hours of travel time}_{ij})^{0.8} \qquad .$$

This expression captures both the pecuniary and time (opportunity) cost of shipping. Hummels and Schaur (2013) estimate that each day in transit is equivalent to an ad-valorem tariff of 0.6-2.1%. Limao and Venables (2001) find that the cost of shipping one ton of freight overland for 1000 miles is about 2% of value, or about 1% per day. This expression generates the resulting target of a loss of 1.6-3.1% in value per day while also incorporating some concavity. Because the transformation from travel time to iceberg cost is necessarily speculative, we checked the robustness of all of our relevant results to alternative calculation of τ_{ij} based on values of ρ between 0.5 and 2.

The calculation of overseas shipping costs requires that we calculate the cost of shipping to the nearest port, and the cost of shipping from that port to an international destination.

[3] The nine ports that handle the largest volume of international trade in 2001 were: Tianjin, Qinhuangdao, Dalian, Shanghai, Lianyungang, Ningbo, Qingdao, Guangzhou, and Shenzhen.

[4] We use ESRI's network analyst for these calculations.

Specifically, to calculate τ_{ix} we use

$$\tau_{ix} = 1.15\tau_{ip} \tag{1}$$

Anderson and van Wincoop (2004) carry out a full accounting of international shipping costs. They conclude that time costs are about 10% (Hummels, 2001) and shipping costs are 1.5% (Limao and Venables, 2001). We treat the cost of shipping from i to the nearest international port p the same as shipping to any other domestic location.

2.5 Measures of access to regional domestic and international markets

With road maps, travel time to port and pairwise iceberg trade costs in hand, we turn to the problem of measuring how the road network affects access to markets. This measurement problem is central to our analysis and raises two main issues. First, we must distinguish between access to international and regional domestic markets. Second, we confront the fact that roads connecting important trading partners are more important than those that do not, but that any measure of domestic access which involves the outcomes of other prefectures gives rise to a structural endogeneity challenge.

Efficiency km of regional roads and travel time to an international port

Our primary measure of 'access to regional domestic markets' is the log 'efficiency kilometers' of highways within the 450 km disk centered on each prefecture city. We assign a weight of one to regular road kilometers and a weight of $\frac{90}{25}$ to limited access highway kilometers.[5] Regional variation in this measure is depicted in Figure 2a, while descriptive statistics appear in table A1. This measure deliberately relies only on the quantity of physical infrastructure and not on regional economic activity. Since we build infrastructure and not 'market access', this eases interpretation.

Our efficiency kilometers measure is based on highways within the 450 km disk centered on each prefecture city. We choose 450 km because evidence indicates most domestic trade occurs over short distances (Hillberry and Hummels, 2003). However, our results are robust to different choices of radius, provided it is small enough to preserve cross-prefecture variation in the measure of efficiency kilometers. We do not have sufficient statistical power to separately estimate effects of infrastructure by distance ring.

We weight limited access highways more heavily in our efficiency kilometers measure for three reasons. First, these are bigger roads that accommodate more people and freight. Second, the particular weights reflect a rough guess at speed of travel along the roads. Third, we would like to compare reduced form regression results with structural counterfactuals. In these structural counterfactuals, we represent 'removing the limited access

[5] More precisely, the 2010 map records three road classes, National Roads, Highways, and Highgrade Highways. We weight the first two classes at unity and the third, the limited access highway class, at 90/25.

highway system' by setting the speed on this network to the same 25 kph that we assign other classes of roads (see section 2.4). A natural reduced form analog to this counterfactual experiment is to recalculate our efficiency road measure weighting all classes of roads equally.

Our primary measure of 'access to international markets' is travel time to the nearest major international port along the 2010 road network. These times are the same as those on which the calculation of τ_{ix} is based in equation (1). Note that better 'access to an international port' is inversely related to travel time to this port, so care is required in the interpretation of regression coefficients. Figure 2b depicts port travel time variation.

Market Potential

Highways to nowhere probably have different impacts than highways connecting potential trading partners, and quantity measures of infrastructure, like efficiency km, will not generally reflect this.

As a first step towards addressing this issue, we construct the following measure of 'market potential', the discounted sum of GDP surrounding each prefecture.

$$MP_i = \sum_j \frac{Y_j}{\tau_{ij}^{\sigma-1}}. \tag{2}$$

Theoretical foundations for this sort of formulation of market potential include Redding and Venables' (2004), Hanson (2005) and Head and Mayer's (2005) adaptations of Fujita, Krugman and Venables' (1999) NEG model. These models feature production of varieties and CES preferences over varieties with elasticity of substitution parameter σ.

This market potential measure has the intuitive property that it weights travel links by the size of demand in each destination j. We considered variants using different calculations of travel time, the shape parameter on the iceberg transport cost, and measures of output. However, reported results use prefecture GDP in 2010, iceberg trade costs calculated on the basis of the 2010 road network and $\sigma = 2$. Figure 2c maps the distribution of market potential.

Market Access

Changing trade costs between any two cities may affect trade flows between other pairs of cities. Neither efficiency km of roads within 450km nor market potential will vary with such indirect effects. This raises the possibility that reduced form estimates based on efficiency km of roads or market potential may not detect important general equilibrium effects of the highway network.

To address this possibility, we adapt the Ricardian and 'New Economic Geography' (NEG) structural models to recover an empirical measure of 'market access'. Ricardian models in Hornbeck and Donaldson (2016), Alder (2015) and Bartelme (2015) and NEG models in Redding (2016) and Balboni (2016) all deliver such a measure. We organize our analysis around a market access measure derived from standard Ricardian foundations (Eaton and Kortum, 2002) that follows closely from Hornbeck and Donaldson (2016), but note that a similarly structured equation arises in NEG models. We use this model to describe trade between the 285 prefectures in our study area plus 'the rest of the world'. Subscripts i and j index prefectures, and for trade flows, i generally indicates product origin and j destination and subscript x indicates rest of the world. Y_i denotes city output or GDP and τ_{ij} is pairwise transport cost as defined above. Finally, θ is the dispersion parameter from Frechet distributed productivity draws, which determines the gains from trade between prefectures, with larger values of θ indicating smaller gains from trade.

We develop this model explicitly in the appendix and derive the system of equations that characterizes equilibrium. For the purposes of our reduced form empirics, only a subset of these equations are relevant, the first is an equation for each city's "market access" MA_i,

$$MA_i = \sum_j \tau_{ij}^{-\theta} \frac{Y_j}{MA_j} + \frac{\tau_{ix}^{-\theta} E}{\sum_j \frac{Y_j}{MA_j} \tau_{jx}^{-\theta}}, i = 1, \ldots, 285, \quad (3)$$

where

$$E = \frac{Y_x}{MA_x} \sum_j \frac{Y_j}{MA_j} \tau_{jx}^{-\theta}$$

is the value of exports.

Equation (3) defines market access with a recursive equation, and given data on Y_i, τ_{ij} and E and calibration of θ, we can solve this system of equations for MA. We refer to the first term in equation (3) as 'domestic market access', the second as 'external market access' and the sum of these components as 'total market access' or just 'market access'. Substituting the definition of E into equation (3) shows that the rest of the world is treated symmetrically with the other 285 trading units, in the sense that 'the rest of the world' is indistinguishable from a large remote domestic unit.

This notion of market access captures three intuitive features of the relationship between trade, output and distance. First, market access is increasing in the income of potential trading partners. Second, it is decreasing in the cost of moving goods between trading partners. Third, market access is decreasing in the extent to which potential trading partners have better access to competing trading partners.

To calculate market access, we solve equation (3) numerically using the observed value of Chinese exports, E, GDP in 2010, Y_j, pairwise transportation costs, τ_{ij}, and set $\theta = 5$ to obtain 285 values of MA_i and MA_x.[6] Figure 2d maps the distribution of market access, figure A1 maps domestic and external components of market access separately, and table A1 gives summary statistics.

[6]We experimented with values of θ ranging from 3 to 10. None of the results we report is sensitive to variation of θ in this range.

Two other equilibrium relationships derived from the model are also of interest. They relate prefecture output and population respectively to market access:

$$\ln Y_i = B_0 + \frac{1+\gamma}{1+\theta\alpha} \ln MA_i + \varepsilon_i,\ i = 1, \ldots, 285, \tag{4}$$

$$\ln N_i = B_1 + (\frac{1+\gamma}{1+\theta\alpha} + \frac{1}{\theta}) \ln MA_i + \varepsilon_i, i = 1, \ldots, 285\ . \tag{5}$$

In the Ricardian context γ is the labor share of output, α is the land share, and B_0, B_1 and ε_i are collections of other structural parameters, city specify productivities, land endowments and residential amenities. More detail is available in appendix equation (12).

In this model wages are simply a share of output so $\ln w_i = \ln\gamma + \ln Y_i - \ln N_i$. Substituting from equations (4) and (5), we see that this has two implications. First, the relationship between prefecture wages and market access is identical to the relationship between per capita GDP and market access. Second, wages and per-capita GDP are decreasing in market access. In table 3, we will see mixed evidence for both of these implications.

Equations (4) and (5) look like a regressions of output or population on market access, with productivity, an unobserved component of the land endowment and the local amenity value forming components of the error term. Given calculated values for market access, such a regression is feasible. However, implementing this equation as a regression raises a number of conceptual and econometric challenges. We discuss these challenges in the context of our econometric framework.

The market access measure described by equation (3) is based on a simple model of Ricardian trade where mobility is frictionless, there is constant returns to scale and perfect competition in production, there is no non-traded housing stock and there are no agglomeration effects. Bartelme (2015) adds a non-traded (housing) sector with Cobb-Douglas utility over this local good and the CES aggregate of the traded good and adds endogenous agglomeration economies. Both extensions imply constant elasticity equilibrium relationships between output and population and market access, just as in (4) and (5). Neither extension changes the definition of market access given in equation (3).

In China, despite massive migration, the hukou system is known to constrain mobility. We show in the appendix that assuming complete immobility preserves the same market access equation and constant elasticity equilibrium relationship between output and market access as we obtain under free mobility. Tombe and Zhu (2015) introduce migration frictions into the Ricardian model, as discussed in the Appendix. This adds a migration friction term to equilibrium equations which we do not have the detailed information about inter-prefecture migration flows to estimate. Instead, we turn to an NEG founded model with migration frictions, which does yield one simple estimation equation.

Following Redding (2016) and Balboni (2016), we relax assumptions of perfect competition, constant returns to scale and free mobility by adopting standard NEG fundamentals. Like our Ricardian model, the consumption side of the NEG model also features CES preferences over varieties. But unlike the Ricardian model, the NEG model has internal increasing returns to scale, with labor as the only factor of production. In addition, the model features monopolistic competition and local housing as an element of consumption. Finally, this model allows for imperfect mobility. Mobility frictions are generated by i.i.d. Fréchet "amenity" draws for each location in which prefecture shift parameters capture variation in the distribution of amenity levels, as in the extended Ricardian framework specified above. These NEG foundations with θ replaced by $1-\sigma$, where σ is the elasticity of substitution in consumption, imply the same expression for market access as in equation (3). They also imply a constant elasticity relationship between wages (rather than population or GDP) and market access. Specifically, with migration costs we have the following equilibrium wage equation:

$$\ln w_i = (\sigma - 1)\ln(\frac{\sigma}{1-\sigma}) - \frac{1}{\sigma}\ln(\sigma F) + \frac{1}{\sigma}\ln MA_i + \frac{(\sigma-1)}{\sigma}\ln T_i, \tag{6}$$

where $\sigma > 1$ is the elasticity of substitution between varieties in consumption and F is the fixed cost of production.

With NEG foundations, wages increase with market access because of increasing returns to scale. The greater the elasticity of substitution between goods, the less demand there is for products to be shipped long distances, thereby reducing the influence of market access on local wages. It is of intrinsic interest to examine the relationship between wages and market access measures. As seen in Equation (6), wages rise with market access while under Ricardian foundations the opposite is true. This would potentially offer an opportunity to discriminate between NEG and Ricardian foundations. However, in table 3, we will see that point estimates of the relationship between wages and market access are not distinguishable from zero. Only when we introduce hierarchy considerations, which are outside both models, will we see more definitive relationships.

To sum up, we derive a constant elasticity relationship between market access and population or GDP in a variety of Ricardian type models and a similar wage equation from NEG models. We will examine these empirically to show that the key initial results we get under our road infrastructure and market potential measures carry over to market access measures derived from structural modeling.

3 Econometric Framework

As we saw in section 2.5, theory implies causal effects of pairwise transportation costs on prefecture GDP, population and wages, and so we primarily consider these outcomes. We are interested in estimating causal relationships between these outcomes and efficiency km of roads, market potential, and market access.

3.1 Empirical model

Denote a measure of access to regional domestic markets by L_{it}, access to international markets by E_{it}, and a prefecture outcome by Y_{it}. The main challenge for the empirical work is that infrastructure measures may be partly determined by some of the same unobservables that predict outcomes of interest.

The following statement of our estimation problem specifies how use of an IV estimator may solve this problem.

$$
\begin{aligned}
y_{it} &= a + \beta L_{it} + \psi E_{it} + X_i \delta + u_{it} && (7)\\
L_{it} &= a_1 + \beta_1 L_{i62} + \psi_1 E_{i62} + X_i \delta_1 + \eta_{it}^1 && (8)\\
E_{it} &= a_2 + \beta_2 L_{i62} + \psi_2 E_{i62} + X_i \delta_2 + \eta_{it}^2 && (9)
\end{aligned}
$$

It is possible that some elements of u_{it} are correlated with L_{it} and E_{it} in equation (7). In the model laid out above, prefecture productivity and amenity value represent unobservables in u_{it}. More productive prefectures may have more resources to build highways. But higher productivity also directly generates greater GDP, population and wages. Other mechanisms not considered in standard economic geography models, including prefecture government competency, might also be important omitted variables.

Incorporating the equations (8) and (9) into estimation resolves such endogeneity concerns as long as our instruments L_{i62} and E_{i62}, which are 1962 counterparts of 2010 infrastructure measures, are uncorrelated with unobservables in u_{it}, conditional on controls X_i. We are careful to use the same instruments and set of control variables X_i across outcomes and predictors. This allows our arguments for the conditional exogeneity of instruments, or that $E[L_{i62}u_{it}] \neq 0$ and $E[E_{i62}u_{it}] \neq 0$, to apply across our full range of estimation results. In order to facilitate coefficient comparisons across predictor, outcome and specification within outcome, we maintain the same instruments for all road and access measures throughout the analysis.

When L_{it} and E_{it} are calculated using only information about roads, we face standard identification concerns about omitted variables that may be correlated with these predictors. When we use $\ln MP_i$ as a measure of L_{it}, two additional concerns arise. First, since $\ln MP_i$ is a function of Y_j for all $i \neq j$, recursive substitution reveals a structural endogeneity problem. [7] Second because $\ln MP_i$is defined in terms of the outcome variable, we effectively create a system with two structural equations. One describes the way that market potential responds to Y_i and the other describes the response of Y_i to market potential. This makes it difficult to evaluate comparative statics. These two problems are standard in spatial lag models. Under parametric assumptions about the nature of the data generating process, established techniques exist to recover the spatial lag parameter

[7] If the market potential measure includes own prefecture output directly, the problem of regressing of Y on itself is transparent. Excluding own prefecture does not resolve the problem. To see this, consider a simple case with two observations, $MP_1 = Y_2/\tau_{12}$ and $MP_2 = Y_1/\tau_{12}$. Substituting into (7) gives $\ln Y_{1t} = a + \beta \ln((a + \beta \ln(Y_1/\tau_{12}) + \psi E_{2t} + X_2\delta + u_{2t})/\tau_{12}) + \psi E_{1t} + X_1\delta + u_{1t}$.

of interest β (Kelejian and Prucha, 2010).[8] However, standard spatial lag estimators are not robust to model mis-specification, an essential attribute of any credible analysis. Our solution is to use an IV estimator constructed using information on the 1962 road network only.

Market access suffers from the same two econometric problems that affect market potential and we deal with them through our IV estimator as well. We note the slight disconnect inherent in a single equation regression implementation of structural models. Structural errors in the model's equilibrium conditions, e.g., equations (4) or (5) are the result of an equilibrium process which does not specify any relationship between the error term and the regressors. Therefore, a regression equation that relies on conditional orthogonality to an instrument introduces a restriction not present in the underlying structural model. Second and related, to date no structural model has incorporated a mechanism to assign roads to places. Unsurprisingly, Baum-Snow et al. (2016) and Duranton and Turner (2012) provide evidence that such assignment depends on the productivity of the two endpoints. Our interest in the structural market access measure is primarily to show that results estimated using our preferred infrastructure-based measures of access to regional domestic and international markets are robust to consideration of heterogeneity in their treatment effects, as captured by the structural model.

3.2 Instrument Validity and First Stages

Prefectures with greater GDP and population are likely to have more resources to build highways, reflecting a reverse causal link from outcomes of interest to highways. Moreover, higher levels of government may have provided better highway links to export nodes for prefectures specialized in export-oriented activities. In short, highway construction is likely to respond to travel and shipping demand. Thus, credible empirical results require exogenous variation across prefectures in the 2010 road network.

We rely on the 1962 road network as a source of quasi-random variation. We use this road network to calculate two instruments. The first is 1962 road kilometers within 450 km of each prefecture but outside the boundaries of the prefecture. The second is the travel time, at 90 kph, along the 1962 road network to the nearest major international port. The rationale for these instruments is based on the idea that 1962 roads were built for other reasons but were upgradeable to modern highways at lower cost than would be required to establish new rights of way. Areas with more vintage roads, however low their quality, had lower costs to build their highway systems. As a result, ceteris paribus, locations with more 1962 roads also had more highways in 2010. We exclude 1962 roads within the prefecture because we are concerned that serially correlated unobservables may predict a prefecture's own 1962 highways and 2010 prefecture outcomes. For example, serially

[8] Gibbons, Overman and Pattacchini (2015) discuss the pitfalls of using these methods. In particular any heterogeneity in β would render all parameter estimates recovered using a standard spatial lag estimator inconsistent.

correlated unobserved components of prefecture productivity may have driven pre-1962 road construction and subsequent growth.

These instruments are only valid if they are strong predictors of 2010 regional and international market access measures and if they are not correlated with unobserved factors that predict outcomes of interest. Therefore, it is important to control for exogenous predictors of GDP and population in 2010 that may be related to the prevalence of roads in 1962. Because 1962 roads were more prevalent in more agriculturally oriented and populous prefectures, we control for 1982 industry mix, education and population throughout our analysis.[9] Because 1962 roads primarily served as connections from agricultural areas to nearby cities, we also control for urbanization with 1982 prefecture city population. We also control for roughness to proxy for agricultural productivity, and for distance to the coast and central city roughness. Finally, provincial capitals have distinct institutional and industrial histories from other prefectures and so we also control for provincial capital status.

Column 1 of table 1 shows the result of regressing the log of 2010 efficiency km of roads within 450 km of prefecture cities on our two instruments and control variables. In addition to being a 'first stage' regression, one can think of this regression equation as a highway supply function. We see a strong relationship between 1962 roads and 2010 roads, conditional on controls, with a significant estimated elasticity of 1.05. Conditional on prefecture area, more populous prefectures had more highways built nearby. The coefficient on prefecture area is negative as expected, with larger prefectures leaving relatively less residual area within which to measure highway length. Interestingly, larger and more manufacturing oriented cities had less highway mileage built nearby, perhaps because manufactures traditionally traveled primarily by rail. Prefectures in the West had less highway length nearby, as is expected given the smaller amount of economic development in these areas. Results are similar when using larger or smaller distance rings than 450 km.[10]

Column 2 of table 1 shows the result of regressing the 2010 road travel time to the nearest international port on the same set of variables. The key predictor in this regression is the 1962 counterpart of dependent variable. This variable has the predicted strong positive relationship, with an estimated elasticity of 0.76. In addition, 10% more 1962 roads within 450 km outside of the origin prefecture reduces port travel time by 3%. Prefectures further from the coast also had longer travel times conditional on the road network and prefecture characteristics, as may be expected. The distance to coast variable is a key control, as areas nearer to the coast tend to be better developed and have better infrastructure.

Columns 1 and 2 of table 1 show that our instruments are strong. These results also confirm our expectation that 1962 regional roads instrument predicts 2010 efficiency km,

[9] 1982 is the first year for which we have information on these variables.

[10] We would like to investigate the relative value of roads nearer and further from a prefecture, e.g., by considering roads within different distance rings of each prefecture. Unfortunately, we do not have statistical power to separate out exogenous variation in road efficiency km for multiple rings simultaneously.

while 1962 travel time to port predicts its modern counterpart.

Remaining columns of table 1 show that our two instruments are also strong predictors of market potential and market access, conditional on controls, though not as strong as of modern roads. This is not surprising. Unlike efficiency km of roads, market potential and market access measures compound variation in roads with variation in prefecture GDP, while the instrument does not include information about GDP. Column 3 shows that the elasticity between market potential and 1962 roads within 450 km is a significant 0.018, with no estimated effect for 1962 simulated port travel time. The estimated elasticities between total market access and 1962 roads and 1962 simulated port travel time are 0.08 and -0.004 respectively, in column 4. When market access is broken out into domestic and external components, regional roads predicts domestic market access while both regional roads and travel time to port predict external market access.

4 Results

4.1 Efficiency km of roads within 450km and travel time to port

Table 2 reports coefficient estimates based on our main regression equation (7), in which 1962 counterparts serve as instruments for 2010 efficiency km of roads within 450 km and travel time to the nearest major international port.

We first consider effects of travel time to a major international port, seen in the second row of table 2. As expected, reducing travel time to a port increases GDP, population and wages. Results in columns 1 and 2 indicate that 10% less travel time to an international port leads to 1.6% higher GDP and 1% higher population. The larger estimated GDP effect leads to -0.07 elasticity of travel time to port on per capita GDP. We also find a positive effect on private sector wages, with this elasticity estimated at -0.04. Because these estimates are conditional on distance to the coast, they are driven by variation in the road network.

The first row of table 2 estimates the effects of regional road capacity. These results are less straightforward than those for travel time to port. The estimated effect of efficiency km of regional roads on GDP and GDP per capita (columns 1 and 4) is not distinguishable from zero, while the effect on private sector wages (column 5) is negative. This suggests the effects of efficiency km of regional roads on productivity is at most small, and possibly negative. Columns 2 and 3 describe the effects of efficiency km of regional roads on population and population growth from 1990-2010. These effects are negative and different from zero at ordinary confidence levels, 10% more road capacity nearby leads to about a 1.2% smaller prefecture population, a strong dispersion effect.

Replicating the results in table 2 without controls (not shown) we find that the relationships between regional roads and both population and GDP are positive, but for population growth (where fixed historical conditions are being differenced out), the coefficient remains negative and significant and little changed at -0.11. Comparing these results with those in table 2 implies that higher GDP and population regions had more roads in 1962 and in 2010, but these locations gained less population than otherwise would have been expected given their other characteristics. Table A2 reports OLS regressions corresponding to table 2. OLS and IV results are qualitatively similar.

The control variables that influence coefficients in columns 1 and 2 the most are 1982 prefecture population and the provincial capital dummy. These two controls have large significant coefficients in table 2 and historical evidence indicates that road infrastructure was historically built to serve agricultural shipments in more populous prefectures and to connect to provincial capitals. However, infrastructure coefficients are not affected much by excluding these controls in the population difference specification in column 3, with resulting coefficients of -0.14 and -0.063 on regional roads and port time respectively.

4.2 Market potential and market access

Table 3 shows baseline estimates of the effects of market potential and market access on the same set of outcomes as table 2. Panel A gives results for market potential. Because the only source of variation in access to international markets is travel time to port, we also include this measure in the market potential regressions. The market potential results mirror those for efficiency km of roads, with exactly the same signs and similar significance levels as in table 2. The much higher absolute magnitudes of coefficients for market potential reflect the small standard deviation of the market potential variable compared to efficiency km of roads (see table A1). We note that, unlike the efficiency km of roads regressions of table 2, there are large differences between OLS and IV market potential results. OLS coefficients on market potential are larger than IV coefficients in all specifications. This is expected. OLS estimates suffer from the standard upward bias that comes with OLS estimation of models with positive spatial lag coefficients.

Panel B of table 3 considers market access effects. Since market access explicitly incorporates international trade, we do not include travel time to port in this regression. We estimate a positive elasticity of market access with respect to GDP of 2.9.

In table 3 panel C we split overall market access into its domestic and external components. Given structural equations with unified market access, coefficients on market access components are predicted to be scaled by the share of that component in total market access. Given summary statistics in table A1, the model predicts about 70% of the total market access effect should be domestic with the remaining 30% external.[11] Evidence in table 3 panel C contravenes this prediction. Domestic market access effects are zero to negative whereas external market access effects are universally positive. Qualitatively, evidence in table 3 panel C is consistent with the more reduced form evidence in table 2.

[11] From equation (3), market access is the sum of a domestic and international component. Decomposing the log of the sum, in which A is the domestic component and B the international component of MA, $\frac{d\ln(A+B)}{dx} = \frac{A}{A+B}\frac{d\ln A}{dx} + \frac{B}{A+B}\frac{d\ln B}{dx}$.

Better access to external markets increases all measures of economic activity and better access to domestic markets decreases these measures.

We note that table 3 poses a number of problems for the models underlying our market access measure. First, on the basis of equation (4) we are able to calculate the coefficient on the MA term in this regression from estimates of the dispersion parameter θ, labor share of output γ and the land share of output α. If, consistent with estimates in the literature we choose $\theta \in [3, 10]$, choose the labor share to be about 0.6 and the land share to about 0.1, then we calculate the coefficient of MA to be in the range $[0.3, 0.5]$. In contrast, at about 3.0 our estimates of this value in table 3 are dramatically larger. Second, an implication of the Ricardian model is that the market access coefficient for wages must be equal to that for per capita output. Comparing table 3, row 3, columns 4 and 5, we see that this relationship certainly does not hold. This seems to rule out the simple formulation of the Ricardian model on which our market access measure is based. Third, in the wage equation derived from NEG foundations, equation (6), we see that the coefficient of MA in such a regression must be positive and inversely related to the elasticity of substitution. In fact, point estimate of this coefficient in table 3, row 3 column 5 are negative, though not distinguishable from zero. Fourth, the opposite signs of domestic and external components of market access in panel C of table 3 seems hard to reconcile with the underlying theory.

In sum, results in table 3 suggest that standard models based on simple formulations of Ricardian or NEG foundations fail to provide a reasonable description of how transportation infrastructure affects economic geography and thus do not provide a basis for estimating how transportation networks affect aggregate economic activity. Given this, we do not utilize market access measures in the main work to follow.

4.3 Main Results: Regional primates, their hinterlands, and the road network

We now show that effects of improved regional road infrastructure on a prefecture are related to the prefecture's regional importance. With improved road access, regional primates gain economic activity at the expense of nearby cities. We show that these effects for primates are not driven by their provincial capital status, absolute population, or centrality in the national highway plan. Their position in the hierarchy of regional cities appears to be their key attribute.

We assign the 26 largest urban centers in 1982 within a 360 minute drive over 1962 roads at 90 kph to be regional primates. We select this six hour cutoff statistically. To select this time, we first estimate a series of regressions analogous to those in table 2 columns 1 and 2 but with the two infrastructure variables interacted with a dummy variable for prefecture primacy, where prefecture primacy is defined on the basis of a candidate driving time radius. Figure 3 shows χ^2 statistics for the joint significance test of whether primacy interactions equal 0 as the driving time radius used to define the regional primate indicator varies between 100 and 600 minutes. When we try to predict prefecture population, the largest χ^2 statistic occurs when this driving radius is 360 minutes, although the value of this test statistic is close to 12 throughout the 340-440 minute range. When we try to predict prefecture GDP, the χ^2 statistic does not vary with driving radius and is everywhere below levels that indicate interactions are statistically significant. To sum up, regional primate interaction effects are most important as determinants of prefecture population when the radius over which 'primacy' is defined is 360 minutes of driving time, and primate status is never important for predicting prefecture GDP. Given this, we organize our analysis around a definition of 'regional primate prefecture' based on the 360 minute driving time radius.

Table 4 panel A reports regressions analogous to those in table 2, but with infrastructure variables interacted with urban primacy. As we saw in table 2, the effects of efficiency km of roads on population, population growth and wages remain negative, but are about are about 50% greater in magnitude for non-primate cities, while the negative effects for GDP larger and the positive effects for GDP per capita smaller, although still not statistically significant. In contrast, regional primates experience significant offsetting positive effects for all outcomes except GDP per capita. Note that the sum of primate and non-primate coefficients is positive for both GDP per person and private sector wages, but this sum is still not distinguishable from zero. It is difficult to make the case, even for regional primates, that efficiency km of roads within 450 km are making important contributions to productivity.

On the other hand, access to international markets affects primate and hinterland prefectures in about the same way. In contrast to efficiency km of roads within 450km, we do not estimate any statistically significant differential effects of port access for regional primates except for private firm wages, where regional primates experience larger effects of port access.

As a robustness check, we also consider the continuous measure of regional primacy status defined earlier. Recall that this indicator is defined for each prefecture by taking the ratio of its 1982 population to that of its regional primate. Panel B of table 4 shows results analogous to those for our regional primate indicator presented in panel A. These results are compelling. The nearer is a prefecture's population to that of its regional primate, the more the negative effects of being a hinterland city are offset. Prefectures that are small relative to their regional primate experience significant negative effects for all scale and productivity measures except per capita GDP. Interaction terms for the continuous primacy variable are positive and highly significant for GDP and population.

Table 4 indicates that regional primates are affected differently by regional roads than are hinterland prefectures. However, evidence on the effects of port connections is more mixed. When primacy is defined as a binary variable, we see positive though insignificant interaction terms on port distance in panel A, with implied lack of port access effects for primates remaining significantly negative for all outcomes except GDP per capita. When defined as a continuous variable, however, these positive interaction terms become marginally significant in the first three columns, resulting in failure to reject that regional primate effects are 0 for all outcomes. Taken together, this suggests that regional primates are less affected by the cost of trucking goods to international markets than hinterland prefectures.

Results in table 4 are based on our preferred measures of transport from table 2. Table A4 in the Appendix shows strong and remarkably similar results to panel A of table 4 for market potential. Model based market access measures in panel B of table A4 show qualitatively similar results but are statistically weaker.

4.4 Robustness of Results to Primacy Definition

Our definition of regional primates was motivated by ideas from central place theory. Here we show that our definition is the one for which heterogeneous effects matter; other definitions of regional importance do not exhibit similar heterogeneous effects. Table 5 reports results analogous to those in table 4 but with a different primacy definition in each panel.

In panel A, we look at nodal cities in a the "5-7" highway plan from the early 1990s. These are cities in which various highways were planned to converge, and thus were viewed as nationally important by the central government at the time. Within our sample there are 38 nodal cities, of which 7 are also regional primates. In panel B, we look at the 29 top 10% population center cities in 1982, of which 7 overlap with our primate definition. In panel C, we look at 24 provincial capitals, of which 7 are also regional primates. (The sets of 7 regional primates that overlap in each of panels A-C are not the same across panels.) In panel D, we just look at the 17 provincial capitals that are not regional primates.

Table 5 presents strong evidence that regional hierarchies matter for regional infrastructure effects, even when accounting for other variables that may be correlated with such primacy. Nodal cities show interaction effects that are all near 0 (panel A). If anything, high population cities are more disadvantaged by an improvement in regional road capacity than other cities (panel B). Only in panel C is there a hint that provincial cities are different from other cities. All differential effects for provincial cities in panel C are positive, though only that in the population difference specification is marginally significant. Panel D shows that these positive interaction effects for regional roads are generated by the handful of regional primates in the group. We find no significant effects of regional roads for provincial capitals that are not primates.

Our primacy definition is motivated by models that think about interregional rather than international trade linkages. It is thus sensible that regional primate results for effects of port connections are less clear than for effects of regional roads. Evidence in table 5 consistently shows relatively large interactions between port travel time and the various measures of regional prefecture importance considered. These interaction coefficients are statistically significant for provincial capitals that are not primates by our definition. This is evidence that nationally important cities have better access to international markets than do other locations, access that depends less on their links through road system. As such, it is not so clear if primacy or some other correlated attribute is driving differential effects of port access.

4.5 Sector-Specific Effects Through the Hierarchy

We expect sectoral differences in responses to a better regional road network. Hinterland producers of traded goods with low land shares, high fixed costs, or that benefit more from agglomeration economies should depart for larger cities or go out of business once the hinterlands become better connected. Traded services (finance, insurance, real estate and business services) and many manufacturing goods have these features. Agriculture has a high land share and so should respond in the opposite way. That is, hinterland areas should become more specialized in agriculture with a better regional road network. Non-traded services should not respond to the regional road network, except through general equilibrium effects on local demand. Conditional on domestic linkages, improved international market linkages may have more complicated effects that depend on aggregate conditions in these different sectors. Using employment data by sector, we verify the expected signs of these responses and measure magnitudes.

Using the same regression specification and primacy definition as in table 4, table 6 estimates the effects of greater regional road capacity and better port access on prefecture employment by industry. The first column shows that estimated effects on total employment, from the 2010 population census, are similar to the population effects reported in table 4 column 3. Subsequent columns decompose these total employment effects into impacts on employment in agriculture, manufacturing, traded services and non-traded services.

In contrast with total employment, the effects of regional roads on agricultural employment are positive for primates and non-primates alike. 10% more roads leads to 4% more agricultural employment. Moreover, access to ports is negatively related to agricultural employment with a 10% greater port travel time leading to 1% more agricultural employment. This reflects substitution with more trade-oriented products. In column 3 we see that, like total employment, manufacturing employment responds positively to roads, but is more sensitive. Negative employment effects for regional non-primates of -0.35 are counteracted by net positive effects for primates of 0.22. While traded services (finance, insurance, real estate and business services) respond like manufacturing to roads, non-traded services have 0 estimated effects of regional roads for primates and non-primates alike. Port access positively affects manufacturing employment in non-regional primates only. Traded services are more greatly affected by port access than non-traded services.

Results in table 6 panel B show qualitatively similar results as in panel A when primacy is defined continuously, with one exception. When primacy is defined continuously, we find that traded and non-traded service employment in regional primates does not benefit from better port access. Effects of both domestic and international road access vary as functions of prefectures' locations in regional hierarchies for employment in all sectors except agriculture.

4.6 Mechanisms Driving Port Access Effects

In this sub-section, we show that the effects of improved port access operate through enhanced trade orientation of the local economy. Using customs data, we calculate the aggregate values of industrial imports and exports into and out of each prefecture in 2010. We relate these measures to the same road and port access measures used throughout the study. Table 7 reports these results, using the same regression specifications as in table 4.

In table 7 we see that the effects of both regional roads and port access on exports and imports are similar. There is almost no impacts on net exports. Thus, effects we find are related to trade orientation of local economies rather than increases in net exports. Estimated responses to regional roads indicate that all else equal, more regional roads lead to lower imports and exports, especially for non-primates. In particular, 10% more regional roads leads to about 20% lower imports and exports in non-primates and 16% less in primates. When viewed in the context of the results in table 4, improved regional road capacity thus increases connections to the domestic economy at the expense of internationally oriented economic activity.

In contrast, better connections to ports increases exports and imports as expected. A 10% reduction in port travel time increases exports by 5.2% in non-primates and 1.5% in primates. The port access estimates from table 4 of -0.31 for non-primates and -0.05 for primates are in line with the table 7 results that non-primates are helped more by road access to ports than are primates. However, the effects must come through some combination of intermediation and productivity effects of better port access.

5 Counterfactual Prefecture Populations Absent Highways

Our final exercise is to examine the cross-sectional distribution of population absent the highway infrastructure built since 1990. We consider the hypothetical reduction of highway speeds to 25 kph, calculate the implied population change for each prefecture, and then adjust each prefecture's population by a constant to equalize initial and final aggregate populations. Since aggregate GDP cannot be assumed constant under counterfactual road networks, we do not consider the corresponding exercise for GDP.

Table 8 shows the results. Columns 1 shows actual minus counterfactual populations that result from setting all highway speeds to 25 kph. In practice, this amounts to giving expressways a weight of 1 rather than 90/25 in the efficiency km calculation. Column 2 shows analogous results from setting port travel speeds to 25 kph. Column 3 shows results of both exercises simultaneously, normalizing the resulting nationwide aggregate population change to 0. The normalization procedure rescales the population of each prefecture by $\frac{\sum N_j^{2010}}{\sum N_j^p}$, where N_j^p is regression predicted population in prefecture j, to result in no change in aggregate national population. Results in columns 1 and 2 are not normalized to sum to 0.

Results in panel A use the non-interacted regression equation from table 2 as a basis. Here, we see that reducing regional highway speeds to 25 kph increases prefecture populations by about half a million on average, with an even balance between primate and non-primate prefectures. However, increasing port travel times has almost exactly an offsetting effect on average prefecture population change. Overall in column 3, primate prefectures have slight average growth associated with reducing road infrastructure and a larger standard deviation than non-primates, mostly because of primates' larger baseline populations.

In contrast are results in panel B which incorporate the primate prefecture interaction. We see in the second row that primates experience population losses because of reduced regional highway speeds. When added to the predicted losses from reducing port access, the empirical model suggests very large population losses for primate prefectures if the expressway system had never been built. In contrast, the empirical model generates small predicted population increases in non-primate prefectures, with the positive effect of reduced regional expressways being substantially offset by losses from reduced access to international ports. Figure 5 provides a visualization of these population results. In Figure 5b it is evident how different the regional primate responses are than those for other cities, something that is missed by the un-interacted empirical model.

If we think of the counterfactual offering up, ceteris paribus, the effects from building the expressway system, we can compare the numbers with true population chnages, normalized to be comparable to column 3. Actual normalized changes are a 337,500 average increase for primates and a 34,000 loss for non-primates. Panel A gives the wrong direction with primates 'predicted' to lose from the construction of expressways and non-primates to gain. In contrast panel B gets the directions correct, with primates predicted to gain 1.2m and non-primates to lose on average 121,000, ceteris paribus. These numbers give larger predicted gains and losses than actually occured, but then building expressways was just one of many changes in this dynamic 20 years for China.

6 Conclusion

The Chinese national highway system has had surprisingly complex effects on the economic geography of China.

Highways that affect access to regional domestic markets, on average, decrease prefecture population and economic activity. Although this seems surprising, it is consistent with canonical intuition about transportation costs: as transportation costs fall, people spread out.

Highways that affect access to international markets, have different effects. Prefectures with better access to an international port are larger, experience greater population growth, produce more output and have higher sector wages. Importantly, better access to international markets is the main feature of the road highway network that leads to higher output per person, a finding that is robust to whether we measure access to international markets with travel time to port or international market access.

These averages reflect heterogeneity in the way that highways affect prefectures at different ranks in the regional hierarchy. Regional primates in the center of a dense regional highway network are larger, grow faster, produce more and have higher private sector wages. Interestingly, the effect of regional highways on GDP per person is too small to distinguish from zero. Regional primates in the center of a dense regional highway network also become relatively specialized in business services and manufacturing, at the expense of agriculture.

Hinterland prefectures in the center of a dense network of highways experience approximately opposite effects. They are smaller, grow more slowly, have less economic output and lower private sector wages. Any effect on output per person is too small to distinguish from zero. These prefectures also become relatively more specialized in agriculture at the expense of manufacturing and services.

Access to international markets affects primate and hinterland prefectures in about the same way, although point estimates suggest that primate prefectures are usually less affected by access to international markets than are their hinterlands.

Finally, and unsurprisingly, improved access to international markets increases the value of imports and exports, although it does not affect the difference between them, i.e., the value of net exports. Better access to regional domestic markets, however, decreases international trade. This suggests that better access to domestic markets leads prefectures to substitute away from foreign goods, and the converse.

These findings suggest a few principles to help to guide transportation policy in China. First, better highways and roads have implications for output per person, only to the extent that they facilitate international trade. All of our other results are easier to reconcile with a shifting of economic activity and population from one place to another. Second, better highways and roads tend to favor regional centers and to encourage the specialization of more hinterland areas into agriculture and more primate areas into manufacturing and services. This suggests that highways construction is at best an uncertain strategy for promoting development in hinterland regions.

With this said, our conclusions are limited in an important way. Our reduced form methodology identifies the way that highway affect one prefecture *relative* to another. To the extent that highways contribute to the growth of all prefectures, this is invisible to our regressions. For the purposes of understanding how population shifts from one region to another this is probably not important. However Chinese real GDP per person increased by about a factor of four during our study period and understanding the role that roads and highways played in this process remains an important question. Purely empirical approaches to this question probably require country level variation in highways and economic activity, and the obstacles to collecting such data and obtaining causal estimates appear formidable. Given this, it seems likely that our understanding of the relationship between transportation infrastructure and the country wide level of economic activity will ultimately rely heavily on theory. Our results also shed some light on the development of such a theory.

Several of our findings suggest that standard models based on simple formulations of Ricardian or NEG foundations probably fail to provide a reasonable description of how transportation infrastructure affects economic geography and thus do not provide a basis for estimating how transportation networks affect aggregate economic activity in China. In particular, the opposite signs of domestic and internal components of market access in panel C of table 3 seem hard to reconcile with the underlying theory, as does the heterogeneity in how access to markets affect primate versus hinterland prefectures. Finally, central to our investigation is the role of quasi-random variation in establishing causal effects. Fundamentally, this reflects the fact that roads and highways are assigned to pairwise links on the basis of the gains from these links. This important relationship is missing from standard models and, to the best of our knowledge, from all extant models of economic geography based on Ricardian or NEG foundations.

With this said, our findings suggest Chinese highways do allow regions to specialize and pursue their comparative advantage. In particular, prefectures where land is abundant, i.e., hinterland prefectures, become more specialized in agriculture, while more centrally located prefectures specialize in manufactured goods for regional consumption. Urban hierarchies appear to be of first order importance to understanding how transportation infrastructure affects economic geography. This suggests that attempts to value transportation infrastructure on the basis of models that do not explicitly deal with the urban hierarchy, the construction of transportation infrastructure and the importance of land endowments should be regarded with suspicion. It also suggests that the development of models with such features should be a fruitful area for further research.

A Derivation of Model Equilibrium Conditions

Here, we derive equation (4) and equation (5) from Ricardian foundations based on Eaton & Kortum (2002) and Donaldson & Hornbeck (2016).

The marginal production cost of a unit of a variety v produced at location i is $\frac{q_i^a w_i^\gamma r^{1-\alpha-\gamma}}{z_i(v)}$, where $z_i(v)$ is productivity, q_i is land rent, w_i is the wage and r is the cost of capital. This Cobb-Douglas form delivers $\gamma Y_i = w_i N_i$ and $\alpha Y_i = q_i L_i$, in which Y is total output, N is labor and L is land. Firms in each location each receive their productivity z_i drawn from a Fréchet distribution with CDF $F_i(z) = 1 - \exp(-T_i z^{-\theta})$.

Consumers shop around for the lowest cost producer of each variety, taking into account the set of iceberg transportation costs τ_{ij} between all pairs of locations. $\tau_{ij}-1$ is the fraction of the value required to ship each unit of exports from i to j . Given the properties of the Fréchet distribution, Eaton and Kortum (2002) demonstrate that the equilibrium value of trade flows between each pair of domestic origin and destination locations is given by

$$X_{ij} = \kappa_1 T_i (q_i^a w_i^\gamma)^{-\theta} \tau_{ij}^{-\theta} \frac{Y_j}{CMA_j}. \tag{10}$$

In (10), Y_j is destination income or GDP, $\kappa_1 = [\Gamma(\frac{\theta+1-\sigma}{\theta})]^{-\theta/(1-\sigma)} r^{-(1-\alpha-\gamma)/\theta}$ where σ is the elasticity of substitution parameter in preferences, and CMA_j denotes 'consumer market access', which summarizes how accessible competing markets are for provision of goods to d . Adding up the value of all flows into China from this expression, we have $I = \kappa_1 T_x (q_x^a w_x^\gamma)^{-\theta} \sum_d \left[\frac{Y_d}{CMA_d} \tau_{xd}^{-\theta}\right]$. In these expressions,

$$\begin{aligned}
CMA_j &\equiv \kappa_1 \sum_i T_i (q_i^a w_i^\gamma)^{-\theta} \tau_{ij}^{-\theta} + \kappa_1 T_x (q_x^a w_x^\gamma)^{-\theta} \tau_{xd}^{-\theta} \\
&= \kappa_1 \sum_i T_i (q_i^a w_i^\gamma)^{-\theta} \tau_{ij}^{-\theta} + \frac{I \tau_{xj}^{-\theta}}{\sum_j \left[\frac{Y_j}{CMA_j} \tau_{xj}^{-\theta}\right]} \\
&= P_j^{-\theta}
\end{aligned}$$

From (10), we see that more productive and lower cost origins ship more everywhere, more is shipped to nearer destinations with lower values of τ_{ij}, to those destinations with more income, and to those destinations with less competition from other locations. If θ is higher, that means less productivity dispersion, so it is less likely that any given origin is going to have a comparative advantage in producing as many varieties. CMA_j is closely related to the price index P_j for location j. In particular, it aggregates the marginal production costs across locations that supply goods to j. Prices are lower, and consumer market access is higher, in locations that are better linked to other productive locations.

Summing over the value of all trade flows from i to j and x, we derive an expression for total income or GDP at i :

$$Y_i = \kappa_1 T_i (q_i^a w_i^\gamma)^{-\theta} \left(\sum_j \tau_{ij}^{-\theta} \frac{Y_j}{CMA_j} + \tau_{ix}^{-\theta} \frac{E}{\sum_i \kappa_1 T_i (q_i^a w_i^\gamma)^{-\theta} \tau_{ix}^{-\theta}} \right) \tag{11}$$

The second term within brackets is derived by setting Chinese exports E equal to the sum of the value of all trade flows to x and can be rewritten as $\tau_{ix}^{-\theta} \frac{Y_x}{CMA_x}$. We see that GDP is decreasing in local production costs and increasing in destinations' GDP. If nearby destinations have greater consumer market access, total income is reduced because of greater nearby export competition. Denoting the term in brackets as 'firm market access' FMA_i , and inverting (11) to substitute for $\kappa_1 T_i (q_i^a w_i^\gamma)^{-\theta}$ within FMA_i, and substituting for $\kappa_1 T_x (q_x^a w_x^\gamma)^{-\theta}$ in CMA_j using aggregate import flows, we have the following equations, which reveal that $FMA_i = CMA_i = MA_i$ if imports equal exports.

$$\begin{aligned} FMA_i &= \sum_j \tau_{ij}^{-\theta} \frac{Y_j}{CMA_j} + \tau_{ix}^{-\theta} \frac{E}{\sum_j \left[\frac{Y_j}{FMA_j} \tau_{jx}^{-\theta} \right]} \\ CMA_j &= \sum_i \tau_{ij}^{-\theta} \frac{Y_i}{FMA_i} + \tau_{xj}^{-\theta} \frac{I}{\sum_i \left[\frac{Y_i}{CMA_i} \tau_{xo}^{-\theta} \right]} \end{aligned}$$

The use of output information on domestic regions married with trade flow information to and from external markets allows us to construct measures of market access that can be decomposed. This is new to the literature.

With free mobility, it must be the case that the real wage is equalized everywhere, or $A_i \frac{w_i}{P_i} = U$ and $w_i = \frac{U}{A_i} MA_i^{-1/\theta}$. Making use of Cobb-Douglas production, $N_i = \frac{\gamma Y_i}{w_i} = \frac{A_i \gamma Y_i}{MA_i^{-1/\theta} U}$. Substituting for q_i and w_i in (11), we derive equilibrium output in each location, which matches (4).

$$\ln Y_i = \frac{1}{1+\theta\alpha} \ln(\kappa_1 T_i) + \frac{\alpha\theta}{1+\theta\alpha} \ln(L_i/\alpha) + \frac{\gamma\theta}{1+\theta\alpha} [\ln A_i - \ln U] + \frac{1+\gamma}{1+\theta\alpha} \ln MA_i \tag{12}$$

Replacing Y_i with $\frac{N_i MA_i^{-1/\theta} U}{A_i \gamma}$ from above, yields the equilibrium population equation

$$\ln N_i = \frac{1}{1+\theta\alpha} \ln(\kappa_1 T_i) - \ln\gamma + \frac{\alpha\theta}{1+\theta\alpha} \ln(L_o/\alpha) + (\frac{\gamma\theta}{1+\theta\alpha} + 1)[\ln A_i - \ln \overline{U}] + (\frac{1+\gamma}{1+\theta\alpha} + \frac{1}{\theta}) \ln MA_i$$

Given data on exports, we can recover the real value of output outside of China $\frac{Y_x}{CMA_x}$ using $E = \frac{Y_x}{CMA_x} \sum_j \kappa_1 T_j (q_j^a w_j^\gamma)^{-\theta} \tau_{jx}^{-\theta} = \frac{Y_x}{CMA_x} \sum_j \tau_{jx}^{-\theta} \frac{Y_j}{MA_j}$. This allows us to determine how E responds under various counterfactual scenarios.

With no labor mobility, it is straightforward to derive the equilibrium relationship between local output and market access,

$$\begin{aligned}\ln Y_i \quad = \quad & \frac{1}{1+\gamma\theta+\alpha\theta}\ln(\kappa_1 T_i) - \frac{\alpha\theta}{1+\gamma\theta+\alpha\theta}\ln(\alpha/L_i) \\ & -\frac{\gamma\theta}{1+\gamma\theta+\alpha\theta}\ln\gamma + \frac{(\gamma+\mu)\theta}{1+\theta\gamma+\theta\alpha}\ln N_i + \frac{1}{1+\gamma\theta+\alpha\theta}\ln MA_i\end{aligned} \tag{13}$$

through the same process, but imposing fixed prefecture population rather than equal utility across locations.

For considering imperfect labor mobility, we adopt a setup similar to that in Tombe and Zhu (2015). Taking 1990 population in each prefecture as exogenous at $\overline{N_i}$, each individual receives an i.i.d. Fréchet multiplicative utility draw from each prefecture with dispersion parameter ρ. The cost of migrating from prefecture j to prefecture i is fraction of utility μ_{ji}. Incorporating this migration friction delivers the following equilibrium relationships between output and market access:

$$\begin{aligned}\ln Y_i \quad = \quad & C + \frac{\theta\gamma}{\rho\theta\alpha+1+2\rho+\gamma\theta}\ln\left(\sum_j \overline{N}_j \frac{\left[\mu_{ji}\right]^{\rho}}{MMA_j}\right) + \frac{\rho\gamma+1+\rho}{\rho\theta\alpha+1+2\rho+\gamma\theta}\ln MA_i \\ & + \frac{\rho\theta\gamma}{\rho\theta\alpha+1+2\rho+\gamma\theta}\ln A_i + \frac{1+\rho}{\rho\theta\alpha+1+2\rho+\gamma\theta}\ln T_i\end{aligned}$$

where $MMA_i = \sum_j \left[\mu_{ij} A_j w_j / p_j\right]^{\rho}$. Note that this relationship is intermediate between the free mobility case in which $\mu_{ji} = 0$ and the no mobility case in which $\mu_{ji} = 1$ unless $i = j$.

References

Alder, S. (2015) 'Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development,' manuscript.

Allen, T., and C. Arkolakis (2014) 'Trade and the Topography of the Spatial Economy,' *Quarterly Journal of Economics*, 129(3): 1085-1140.

Alvarez, F., F. Buera and R.E. Lucas (2013) 'Idea Flows, Economic Growth, and Trade,' NBER Working Paper 19667.

Anderson, J. E. and E. van Wincoop (2004) 'Trade Costs,' *Journal of Economic Literature*, 42(3): 691-751.

Au, C. C. and J. V. Henderson (2006) 'Are Chinese Cities Too Small?' *Review of Economic Studies*, 73: 549-576.

Balboni, C. (2016) 'Living on the Edge: Infrastructure Investments and the Persistence of Coastal Cities,' manuscript.

Banerjee, A., E. Duflo and N. Qian (2012) 'On the Road: Transportation Infrastructure and Economic Development,' NBER Working Paper 17897.

Bartelme, D. (2015) 'Trade Costs and Economic Geography: Evidence from the U.S.,' manuscript.

Bartik, T. (1991) 'Who Benefits from State and Local Economic Development Policies?' Kalamazoo (MI):W.E. Upjohn Institute for Employment Research.

Baum-Snow, N. (2007) 'Did Highways Cause Suburbanization?' *Quarterly Journal of Economics*, 122(2): 775–805.

Baum-Snow, N. (2017) 'Urban Transport Expansions, Employment Decentralization, and the Spatial Scope of Agglomeration Economies,' manuscript.

Bird, J. and S. Straub (2015) 'The Brasilia Experiment:Road Access and the Spatial Pattern of Long-Term Local Development in Brazil,' manuscript.

Bosker, M., U. Deichman and M. Roberts (2015) 'Hukou and Highways: The Impact of China's Spatial Development Policies on Urbanization and Regional Inequality,' World Bank Policy Research Working Paper 7350.

Branstetter, L. and N. Lardy (2008) 'China'S Embrace of Globalization,' in 'China's great economic transition,' L. Brandt and T. G. Rawski (eds) Cambridge University Press.

Buera, F. and E. Oberfeld (2014) 'The Global Diffusion of Ideas,' manuscript.

Card, D. (2001) 'Immigrant Inflows, Native Outflows and the Local Labor Market Impacts of Higher Immigration,' *Journal of Labor Economics*, 19(1): 22-64.

Chan, K. W. (2001) 'Recent Migration in China: Patterns, Trends, and Policies,' *Asian Perspectives*, 25(4): 127-155.

Chan, K. W. (2005) 'Migration and Small Town Development: Some Notes,' World Bank Workshop, Beijing.

Christaller, W. (1933) 'Central Places in Southern Germany,' Prentice-Hall, Englewood Cliffs, NJ.

Combes, P-P., G. Duranton and L. Gobillon (2012) 'The Costs of Agglomeration: Land Prices in French Cities,' manuscript.

Desmet, K., D. K. Nagy and E. Rossi-Hansberg (2015) 'The Geography of Development: Evaluating Migration Restrictions with Coastal Flooding,' manuscript.

Donaldson, D. (2015) 'Railroads of the Raj: Estimating the Impact of Transportation Infrastructure,' *American Economic Review*, forthcoming.

Donaldson, D. and R. Hornbeck (2016) 'Railroads and American Economic Growth: A 'Market Access' Approach,' *Quarterly Journal of Economics*, 131(2): 799-858.

Duranton, G., P. M. Morrow, and M. A. Turner (2014) 'Roads and Trade: Evidence from U.S. Cities,' *Review of Economic Studies*, 81(2): 681-724.

Duranton, G. and M. A. Turner (2011) 'The Fundamental Law of Road Congestion: Evidence from U.S. Cities,' *American Economic Review*, 101(6): 2616-2652.

Duranton, G. and M. A. Turner (2012) 'Urban Growth and Transportation,' *Review of Economic Studies*, 79(4): 1407-1440.

Duranton, G. and D. Puga (2004) 'The Micro-Foundations of Agglomeration Economies,' Handbook of Regional and Urban Economics.

Eaton, J. and S. Kortum (2002) 'Technology, Geography and Trade,' *Econometrica*, 70(5): 1741-1779.

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

Fajgelbaum, P. and S. Redding (2014) 'External Integration, Structural Transformation and Economic Development: Evidence from Argentina 1870-1914,' manuscript.

Fajgelbaum, P. and A. K. Cosar (2016) ‘Internal Geography, International Trade, and Regional Specialization,’ *Review of Economic Studies: Microeconomics*, 8(1): 24-56.

Fujita, M., P. Krugman and T. Mori (1999) ‘On the Evolution of Hierarchical Urban Systems,’ *European Economic Review*, 43(2): 209-251.

Fujita, M., P. Krugman and A. J. Venables (1999) The Spatial Economy, MIT Press.

Fujita, M. and H. Ogawa (1982) ‘Multiple Equilibria and Structural Transition of Non-Monocentric Urban Configurations,’ *Regional Science and Urban Economics*, 12: 161-196.

Ghani, E., A. G. Goswami, and W. R. Kerr (2016) ‘Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing,’ *The Economic Journal*, 591(126): 317-357.

Gibbons, S., H. G. Overman and E. Patacchini (2014) ‘Spatial Methods,’ in Handbook of Regional and Urban Economics Vol. 5, Duranton, G., J. V. Henderson and W. Strange Eds., Elsevier.

Hanson, G. (2005) ‘Market Potential, Increasing Returns, and Geographic Concentration,’ *Journal of International Economics*, 67(2005): 1-24.

Head, K. and T. Mayer (2004) ‘Market Potential and the Location of Japanese Investment in the European Union,’ *Review of Economics and Statistics*.

Henderson, J. V. (1974) ‘The Sizes and Types of Cities,’ *American Economic Review*, 64(4): 640-656.

Henderson, J. V., A. Kuncoro and D. Nasution (1996) ‘The Dynamics of Jabotabek Development,’ *Bulletin of Indonesian Economic Studies*.

Henderson, J. V., A. Storeygard and D. Weil (2012) ‘Measuring Economic Growth from Outer Space,’ *American Economic Review* 102(2):994-1028.

Hillberry, R. and D. Hummels (2003) ‘Intra-National Home Bias:Some Explanations,’ *Review of Economics and Statistics*, 85: 1089-1092.

Hummels, D. (2001) ‘Time as a Trade Barrier,’ manuscript.

Hummels, D. and G. Schaur (2013) ‘Time as a Trade Barrier,’ *American Economic Review*, 103: 2935-59.

Kelejian, H. and I. R. Prucha (2010) ‘Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances,’ *Journal of Econometrics*, 157:1.

Krugman, P. (1991) 'Increasing Returns and Economic Geography,' *Journal of Political Economy*, 99:3, 483-499.

Lee, K. S. (1982) 'A Model of Intra-Industry Employment Location: An Application to Bogota Columbia,' *Journal of Urban Economics*, 12: 263-279.

Lee K. S. and S. C. Choe (1990) 'Changing Location Patterns of Industries and Urban Decentralization Policies in Korea,' in J. K. Kwon, Editor, Korean Economic Development, Greenwood, New York.

Lewis, E. (2011) 'Immigration, Skill Mix, and Capital Skill Complementarity,' *Quarterly Journal of Economics*, 126(2): 1029-1069.

Limao, N. and A. J. Venables (2001) 'Infrastructure, Geographical Disadvantage, Transport Costs, and Trade,' *World Bank Economic Review*, 15(3): 451-479.

Lucas, R .E. (1988) 'On the Mechanics of Economic Development,' *Journal of Monetary Economics*, 22: 3-42.

Lyons, T. P. (1985) 'Transportation in Chinese Development, 1952-1982,' *Journal of Developing Areas*, 19: 305-328.

Michaels, G., S. Rauch and S. Redding (2007) 'Urbanization and Structural Transformation,' *Quarterly Journal of Economics*, 127(2): 535-586.

Redding, S. and A. J. Venables (2004) 'Economic Geography and International Inequality,' *Journal of International Economics*, 62: 53-82.

Redding S. J. and D. M. Sturm (2008) 'The costs of remoteness: Evidence from German division and reunification,' *American Economic Review*, 98(5): 1766-1797.

Redding, S. (2016) 'Goods Trade, Factor Mobility and Welfare,' *Journal of International Economics*, 101: 148-167.

Roberts, M., U. Deichmann, B. Fingleton and T. Shi (2012) 'Evaluating China's road to prosperity: A new economic geography approach,' *Regional Science and Urban Economics*, 42(4): 580-594.

Rosenthal, S. and W. Strange (2004) 'Evidence on the Nature and Sources of Agglomeration Economies,' in Handbook of regional and urban economics, 4: 2119-2171.

Rossi-Hansberg, E., and M. L. Wright (2007) 'Urban Structure and Growth,' *Review of Economic Studies*, 74(2): 597-624.

Sotelo, S. (2015) 'Domestic Trade Frictions and Agriculture,' manuscript.

Storeygard, A. (2016) ‘Farther on Down the Road: Transport Costs, Trade and Urban Growth In Sub-Saharan Africa,’ *Review of Economic Studies*, 83(3): 1263-1295.

Tabuchi, T. and Thisse, J. (2011) ‘A New Economic Geography Model of Central Places,’ *Journal of Urban Economics*, 69(2): 240-252.

Tombe T. and X. Zhu (2015) ‘Trade, Migration, and Productivity: A Quantitative Analysis of China,’ University of Toronto Working Paper No. 542.

Topalova, P., and Khandelwal, A. (2011) ‘Trade Liberalization and Firm Productivity: The Case of India,’ *Review of Economics and Statistics*, 93(3): 995-1009.

Yang Y. (2016) ‘Transport Infrastructure, City Productivity Growth and Industry Reallocation: Evidence from China,’ manuscript.

Table 1: First Stage Regressions

	Log 2010 Road Efficiency Units within 450km (1)	Log 2010 Time to Nearest Port (2)	Log 2010 Market Potential Gravity (3)	Log 2010 Market Access (4)	Log 2010 Domestic Market Access (5)	Log 2010 External Market Access (6)
Instruments						
Log 1962 Roads Within	1.05***	-0.26**	0.018***	0.080***	0.087***	0.058***
450km, Excluding Own Pref	(0.038)	(0.13)	(0.0017)	(0.0076)	(0.0083)	(0.0080)
Log 1962 Minimum Port Travel	-0.024***	0.76***	-0.00036	-0.0038***	-0.0017	-0.010***
Time Given Road Upgrades	(0.0080)	(0.061)	(0.00030)	(0.0013)	(0.0015)	(0.0014)
Controls						
Log Prefecture Area, 2005	-0.052***	-0.053	-0.0029***	-0.012***	-0.015***	-0.0040
	(0.019)	(0.054)	(0.00078)	(0.0033)	(0.0038)	(0.0036)
Log Central City Area, 1990	0.0055	0.031	-0.000089	-0.00070	-0.00050	-0.0012
	(0.012)	(0.051)	(0.00048)	(0.0020)	(0.0023)	(0.0020)
Log Central City Population,	-0.026*	-0.0076	-0.0012**	-0.0051**	-0.0059**	-0.0028
1982	(0.015)	(0.071)	(0.00058)	(0.0024)	(0.0028)	(0.0027)
Log Central City Roughness	-0.0060	0.041	0.00011	0.00060	0.00062	0.00053
	(0.0097)	(0.050)	(0.00036)	(0.0015)	(0.0017)	(0.0017)
Log Prefecture roughness	-0.019**	-0.040	-0.00044	-0.0020	-0.0020	-0.0021*
	(0.0093)	(0.036)	(0.00031)	(0.0013)	(0.0015)	(0.0012)
Provincial Capital	0.066*	0.048	0.0013	0.0038	0.0060	-0.0022
	(0.038)	(0.12)	(0.0013)	(0.0054)	(0.0064)	(0.0052)
Log Prefecture Population,	0.071***	0.017	0.0033***	0.014***	0.016***	0.0055
1982	(0.023)	(0.081)	(0.00087)	(0.0039)	(0.0042)	(0.0043)
Share Prefecture Population	-0.78**	-1.27	-0.013	-0.035	-0.061	0.040
with High School, 1982	(0.32)	(0.98)	(0.010)	(0.045)	(0.050)	(0.044)
Share Prefecture Population	-0.25	-0.52	0.0016	0.014	0.0070	0.036*
in Manufacturing, 1982	(0.16)	(0.58)	(0.0047)	(0.019)	(0.023)	(0.019)
Log km to Coast	0.0030	0.055*	-0.00053**	-0.0037***	-0.0027**	-0.0067***
	(0.0068)	(0.028)	(0.00026)	(0.0012)	(0.0013)	(0.0017)
West Region	-0.25***	0.054	-0.0045***	-0.031***	-0.022***	-0.058***
	(0.031)	(0.087)	(0.0011)	(0.0050)	(0.0054)	(0.0057)
East Region	-0.014	-0.17	0.00091	0.013***	0.0048	0.038***
	(0.023)	(0.11)	(0.00080)	(0.0034)	(0.0039)	(0.0038)
Constant	1.03***	3.82**	12.8***	5.80***	5.43***	4.68***
	(0.37)	(1.53)	(0.017)	(0.080)	(0.085)	(0.086)
R-squared	0.90	0.88	0.74	0.81	0.75	0.88

Notes: Each regression has 285 observations. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 2: Baseline Infrastructure Regressions

	Log Prefecture GDP 2010 (1)	Log Prefecture Pop 2010 (2)	Prefecture Pop Gr. Rate 1990-2010 (3)	Log Prefecture GDP PerCap 2010 (4)	Log Private Firm Wage 2007 (5)
Infrastructure Variables					
Log 2010 Road Efficiency	-0.032	-0.12**	-0.13***	0.091	-0.11*
Units within 450 km	(0.13)	(0.059)	(0.045)	(0.11)	(0.061)
Log 2010 Minimum Port	-0.16**	-0.098*	-0.068**	-0.057**	-0.042**
Travel Time	(0.066)	(0.052)	(0.028)	(0.026)	(0.018)
Controls					
Log Prefecture Area, 2005	-0.041	-0.059**	-0.053**	0.018	-0.065**
	(0.061)	(0.029)	(0.026)	(0.052)	(0.031)
Log Central City Area, 1990	-0.10**	-0.032	-0.024	-0.069*	0.023
	(0.049)	(0.026)	(0.016)	(0.038)	(0.019)
Log Central City Population,	0.11**	0.023	0.024	0.090*	0.020
1982	(0.054)	(0.025)	(0.018)	(0.046)	(0.028)
Log Central City Roughness	-0.050	0.0013	0.0020	-0.051*	0.037**
	(0.033)	(0.014)	(0.010)	(0.027)	(0.015)
Log Prefecture roughness	-0.022	0.00026	0.0031	-0.023	0.013
	(0.028)	(0.012)	(0.0094)	(0.022)	(0.014)
Provincial Capital	0.65***	0.35***	0.26***	0.29***	0.16**
	(0.11)	(0.051)	(0.038)	(0.086)	(0.069)
Log Prefecture Population,	0.56***	0.83***	-0.094***	-0.27***	0.041
1982	(0.090)	(0.053)	(0.032)	(0.066)	(0.040)
Share Prefecture Population,	0.51	-0.19	-0.33	0.70	-0.58
with High School, 1982	(0.92)	(0.42)	(0.34)	(0.70)	(0.55)
Share Prefecture Population,	1.98***	-0.38	-0.024	2.36***	0.55**
in Manufacturing, 1982	(0.54)	(0.37)	(0.22)	(0.34)	(0.23)
Log km to Coast	-0.021	-0.0087	-0.0046	-0.012	-0.014
	(0.034)	(0.013)	(0.011)	(0.028)	(0.013)
West Region	-0.083	-0.0065	-0.012	-0.076	0.0077
	(0.10)	(0.043)	(0.035)	(0.086)	(0.053)
East Region	0.16*	-0.045	-0.028	0.20***	0.086*
	(0.083)	(0.045)	(0.030)	(0.064)	(0.045)
Constant	-0.59	5.25***	3.69***	-5.83***	10.6***
	(2.03)	(1.40)	(0.83)	(1.28)	(0.78)

Notes: Regressions in columns 1-4 have 285 observations and that in column 5 has 283 observations. First stage regressions are in Table 1. Kleibergen-Paap first stage F statistics are 236 in 1-4 and 237 in 5. Robust standard errors are in parentheses.

Table 3: Effects of Market Potential and Market Access

	Log Prefecture GDP 2010 (1)	Log Prefecture Pop 2010 (2)	Prefecture Pop Gr. Rate 1990-2010 (3)	Log Prefecture GDP PerCap 2010 (4)	Log Private Firm Wage 2007 (5)
Panel A: Market Potential Gravity Regressions					
Log 2010 Market Potential	-1.88	-7.25**	-7.54***	5.37	-6.66*
Gravity	(7.67)	(3.57)	(2.83)	(6.27)	(3.62)
Log 2010 Minimum Port	-0.15**	-0.098*	-0.067**	-0.057**	-0.042**
Travel Time	(0.066)	(0.052)	(0.029)	(0.026)	(0.018)
Panel B: Market Access Regressions					
log 2010 Market Access	2.88*	0.53	-0.17	2.34*	-0.52
	(1.60)	(0.92)	(0.61)	(1.21)	(0.70)
Panel C: Market Access Regressions, Domestic and External					
log 2010 Domestic	-8.58*	-6.63**	-5.08***	-1.95	-3.56**
Market Access	(4.48)	(3.34)	(1.84)	(2.20)	(1.42)
log 2010 External	13.0**	8.15*	5.61**	4.82**	3.47**
Market Access	(5.57)	(4.49)	(2.41)	(2.11)	(1.49)

Notes: Each regression has the same set of control variables as in Table 2. Regressions in columns 1-4 have 285 observations and that in column 5 has 283 observations. First stage F-statistics are 56.3 in Panel A, 69.2 in Panel B and 21.3 in Panel C. First stage regressions are in Table 1. Robust standard errors are in parentheses.

Table 4: Infrastructure Effects with Primate Prefecture Interactions

	Log Prefecture GDP 2010	Log Prefecture Pop 2010	Prefecture Pop Gr. Rate 1990-2010	Log Prefecture GDP PerCap 2010	Log Private Firm Wage 2007
	(1)	(2)	(3)	(4)	(5)
Panel A: Binary Measure of Regional Prefecture Primacy					
Log 2010 Road Efficiency	-0.13	-0.17**	-0.17***	0.042	-0.16**
Units within 450 km	(0.14)	(0.071)	(0.051)	(0.11)	(0.065)
X Primate Prefecture	0.44**	0.28***	0.24***	0.16	0.25**
	(0.18)	(0.089)	(0.072)	(0.15)	(0.12)
Log 2010 Minimum Port	-0.18**	-0.11*	-0.074**	-0.068**	-0.033*
Travel Time	(0.075)	(0.061)	(0.033)	(0.028)	(0.020)
X Primate Prefecture	0.079	0.033	0.010	0.046	-0.054*
	(0.076)	(0.047)	(0.028)	(0.049)	(0.030)
Primate Prefecture	-5.13**	-3.07***	-2.60***	-2.07	-2.38*
	(2.20)	(1.13)	(0.85)	(1.75)	(1.34)
Panel B: Continuous Measure of Regional Prefecture Primacy					
Log 2010 Road Efficiency	-0.34*	-0.34***	-0.29***	-0.002	-0.27***
Units within 450 km	(0.18)	(0.10)	(0.073)	(0.14)	(0.093)
X Primate Prefecture	0.53**	0.45***	0.34***	0.080	0.20
	(0.23)	(0.12)	(0.091)	(0.19)	(0.12)
Log 2010 Minimum Port	-0.31**	-0.23**	-0.15***	-0.077*	-0.033
Travel Time	(0.13)	(0.11)	(0.054)	(0.041)	(0.034)
X Primate Prefecture	0.26*	0.22*	0.13**	0.044	-0.010
	(0.15)	(0.12)	(0.059)	(0.060)	(0.040)
Primate Prefecture	-7.35***	-5.99***	-4.31***	-1.36	-2.23*
	(2.83)	(1.80)	(1.13)	(2.09)	(1.33)

Notes: Each regression has the same set of control variables as in Table 2. Regressions in columns 1-4 have 285 observations and that in column 5 has 283 observations. The Kleibergen-Paap first stage F statistic is 157 for each regression in Panel A and 147 in Panel B. Results for total prefecture employment are similar as those for population. Robust standard errors are in parentheses.

Table 5: Effects Using Alternative Definitions of Primacy

	Log Prefecture GDP 2010	Log Prefecture Pop 2010	Prefecture Pop Gr. Rate 1990-2010	Log Prefecture GDP PerCap 2010	Log Private Firm Wage 2007
	(1)	(2)	(3)	(4)	(5)
Panel A: Primacy as Nodal Prefecture in the 5-7 Road Plan					
Log 2010 Road Efficiency	-0.059	-0.13**	-0.14***	0.072	-0.13**
Units within 450 km	(0.12)	(0.057)	(0.045)	(0.10)	(0.057)
X Nodal Prefecture	0.0022	0.014	-0.0065	-0.011	-0.010
	(0.048)	(0.032)	(0.021)	(0.024)	(0.013)
Log 2010 Minimum Port	-0.19**	-0.094**	-0.087**	-0.093*	-0.051**
Travel Time	(0.083)	(0.046)	(0.036)	(0.048)	(0.026)
X Nodal Prefecture	0.095	0.021	0.045	0.074	0.025
	(0.092)	(0.056)	(0.039)	(0.052)	(0.027)
Panel B: Primacy as Top 10% of 1982 Center City Populations					
Log 2010 Road Efficiency	-0.082	-0.16**	-0.15***	0.079	-0.13**
Units within 450 km	(0.13)	(0.071)	(0.049)	(0.11)	(0.057)
X Large Population City	-0.045	-0.044	-0.022	-0.00049	-0.012
	(0.057)	(0.043)	(0.024)	(0.023)	(0.015)
Log 2010 Minimum Port	-0.22*	-0.15	-0.095*	-0.067*	-0.053**
Travel Time	(0.12)	(0.095)	(0.052)	(0.038)	(0.026)
X Large Population City	0.15	0.11	0.062	0.041	0.044
	(0.10)	(0.079)	(0.044)	(0.042)	(0.028)
Panel C: Primacy as Provincial Capitals					
Log 2010 Road Efficiency	-0.080	-0.14**	-0.14***	0.062	-0.12*
Units within 450 km	(0.13)	(0.065)	(0.048)	(0.11)	(0.063)
X Provincial Capital	0.25	0.11	0.12*	0.14	0.098
	(0.19)	(0.079)	(0.069)	(0.17)	(0.13)
Log 2010 Minimum Port	-0.20**	-0.11*	-0.077**	-0.088***	-0.040*
Travel Time	(0.081)	(0.067)	(0.036)	(0.029)	(0.023)
X Provincial Capital	0.15**	0.051	0.028	0.095**	-0.0057
	(0.070)	(0.049)	(0.028)	(0.040)	(0.022)
Panel D: Primacy as non Rank 1 Provincial Capitals					
Log 2010 Road Efficiency	-0.022	-0.11*	-0.12***	0.090	-0.13**
Units within 450 km	(0.13)	(0.058)	(0.045)	(0.11)	(0.062)
X Non Rank 1	0.056	0.087	0.092	-0.031	0.074
Provincial Capital	(0.23)	(0.069)	(0.066)	(0.21)	(0.14)
Log 2010 Minimum Port	-0.17**	-0.11*	-0.076**	-0.063**	-0.039**
Travel Time	(0.070)	(0.057)	(0.031)	(0.027)	(0.020)
X Non Rank 1	0.15***	0.077**	0.056***	0.074**	0.016
Provincial Capital	(0.053)	(0.034)	(0.021)	(0.036)	(0.022)

Notes: First stage F-statistics are 8.4 in Panel A, 25.1 in Panel B, 34.9 in Panel C and 58.5 in Panel D. Full interactions of any of these three alternative primacy definitions with largest city within 300 minute drive yields first stage F-statistics that are too small for resulting regression results to be informative.

Table 6: Effects on log Employment by Industry

	Total	Agric.	Manuf.	FIRE & Bus. Svc.	Other Services
	(1)	(2)	(3)	(4)	(5)
Panel A: Binary Measure of Regional Prefecture Primacy					
Log 2010 Road Efficiency	-0.19**	0.38***	-0.35*	-0.29**	-0.043
Units within 450 km	(0.089)	(0.11)	(0.19)	(0.14)	(0.072)
X Primate Prefecture	0.34***	0.22	0.57***	0.41**	0.16
	(0.097)	(0.18)	(0.22)	(0.19)	(0.100)
Log 2010 Minimum Port	-0.13*	0.087**	-0.24***	-0.22**	-0.11**
Travel Time	(0.071)	(0.038)	(0.089)	(0.094)	(0.056)
X Primate Prefecture	0.020	0.081	0.21**	0.10	0.054
	(0.056)	(0.065)	(0.096)	(0.093)	(0.053)
Panel B: Continuous Measure of Regional Prefecture Primacy					
Log 2010 Road Efficiency	-0.41***	0.47***	-0.69***	-0.55***	-0.17
Units within 450 km	(0.13)	(0.15)	(0.24)	(0.19)	(0.11)
X Primate Prefecture	0.54***	0.012	0.89***	0.58**	0.34**
	(0.15)	(0.21)	(0.31)	(0.25)	(0.14)
Log 2010 Minimum Port	-0.27**	0.14**	-0.44***	-0.43***	-0.24**
Travel Time	(0.12)	(0.066)	(0.15)	(0.15)	(0.094)
X Primate Prefecture	0.27**	-0.073	0.43***	0.40**	0.23**
	(0.14)	(0.092)	(0.16)	(0.17)	(0.10)

Notes: Each regression has the same set of control variables as in Table 2, with 285 obervations. The First stage F-statistic is 157 for each regression in Panel A and 285 for each regression in Panel B.

Table 7: Highways and External Trade

	log Exports	log Imports	log (Exports/Imports)	log(Exports-Imports)
Log 2010 Road Efficiency	-2.07***	-1.95***	-0.14	-8.04
Units within 450 km	(0.38)	(0.48)	(0.37)	(4.97)
X Primate Prefecture	0.46	0.44	0.14	5.76
	(0.54)	(0.64)	(0.54)	(7.79)
Log 2010 Minimum Port	-0.52***	-0.48***	-0.041	-1.30
Travel Time	(0.14)	(0.14)	(0.050)	(1.50)
X Primate Prefecture	0.37**	0.35**	-0.0076	0.96
	(0.15)	(0.16)	(0.080)	(2.84)
Primate Prefecture	-7.05	-6.40	-1.91	-76.7
	(5.96)	(7.05)	(5.84)	(87.9)
Observations	275	273	273	275

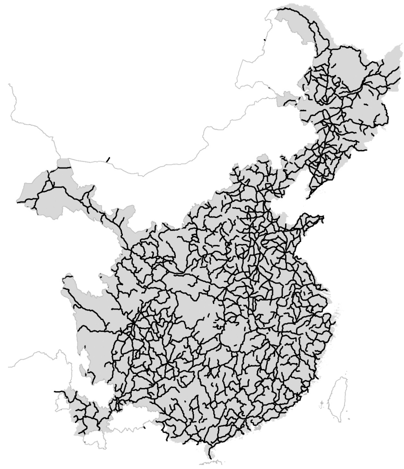
Notes: Specifications are identical to those in Table 4, with the RMB value of prefecture industrial exports and imports used to build dependent variables. The first stage F-statistic is 155 in columns 1 and 4 and 151 in columns 2 and 3.

Table 8: Impacts of Downgrading Expressways on Population
Population Gains Associated with Going from 2010 to 1990 Roads Infrastructure
Prefecture Means with Standard Deviations in Parentheses

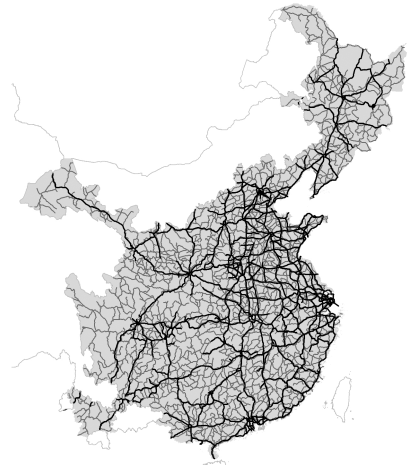
	Highways become 25 kph	Port travel time at 25 kph	Both
	(1)	(2)	(3)
Panel A: Non-Interacted Regression Specification			
Full Sample	514,445	-503,058	0
	(431,944)	(397,052)	(376,019)
Primate Prefectures Only	969,211	-804,382	161,471
	(918,626)	(719,867)	(944,778)
Non-Primate Prefectures Only	468,793	-472,809	-16,209
	(317,291)	(336,534)	(257,399)
Panel B: Primate City Interacted Regression Specification			
Full Sample	531,676	-538,052	0
	(624,731)	(396,773)	(570,695)
Primate Prefectures Only	-748,171	-640,686	-1,206,162
	(694,986)	(572,957)	(947,008)
Non-Primate Prefectures Only	660,154	-527,749	121,082
	(448,526)	(374,587)	(332,091)

Notes: Counterfactuals are calculated based on regression specifications reported in Table 2 Column 2 and Table 4 Column 2. Each entry in Columns 1 and 2 shows the average 2010 prefecture population net of the indicated roads effect minus 2010 prefecture population. Each entry in Column 3 shows the average 2010 prefecture population net of all road effects scaled to sum to 2010 prefecture population minus 2010 prefecture population.

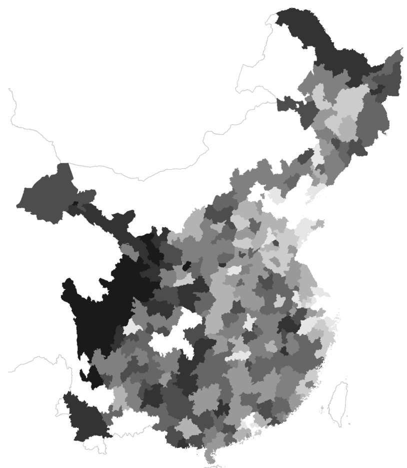
Figure 1: Geographic distributions of data. In all panels but (c) lighter colors indicate larger values. Highlighted prefectures in panel (e) are regional centers.



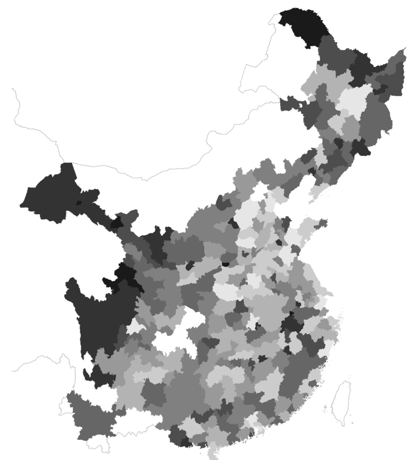
(a) 1962 road network



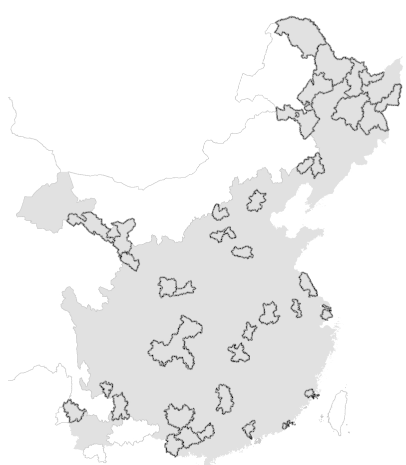
(b) 2010 road network



(c) 2010 GDP

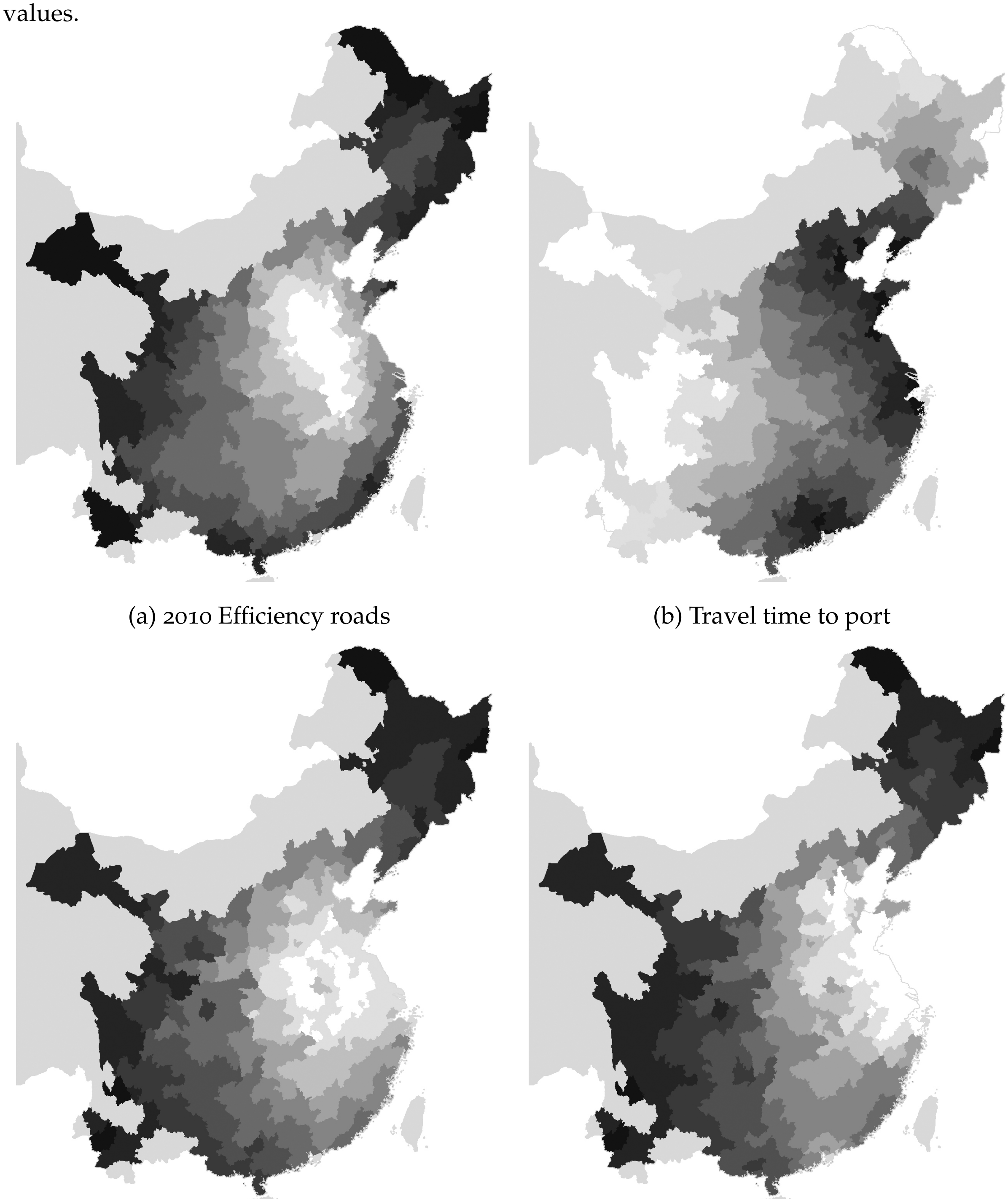


(d) 2010 population



(e) Regional centers

Figure 2: Geographic distributions of road measures. In all panels lighter colors indicate larger values.



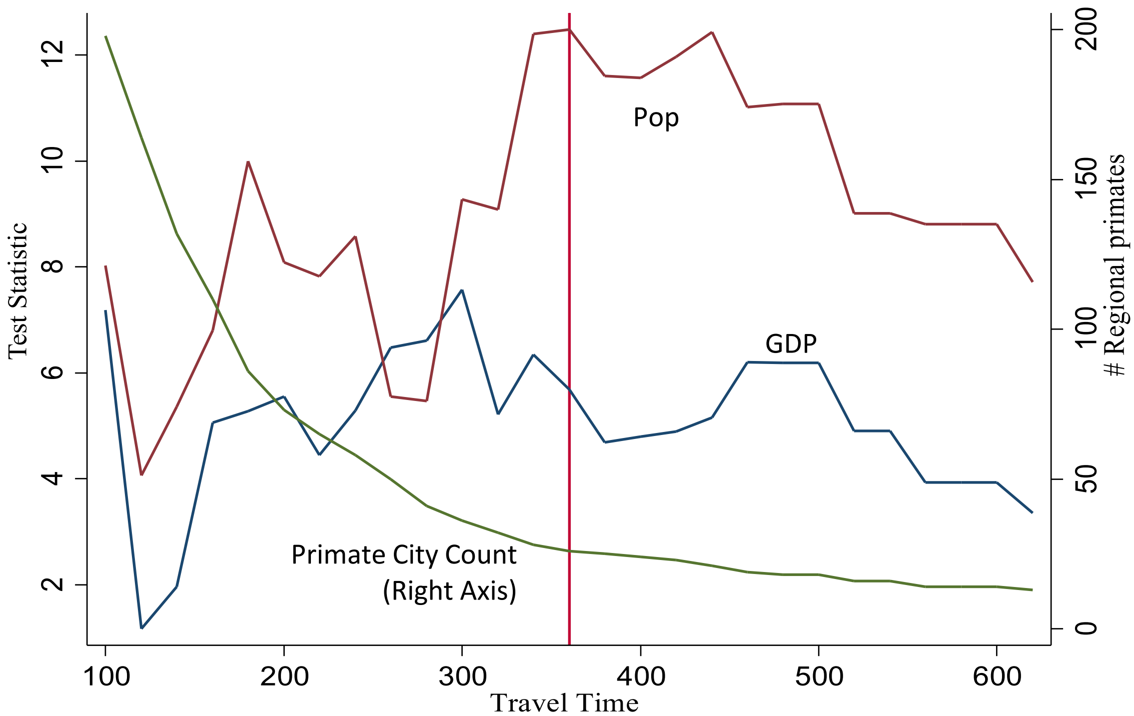
(a) 2010 Efficiency roads

(b) Travel time to port

(c) Market Potential

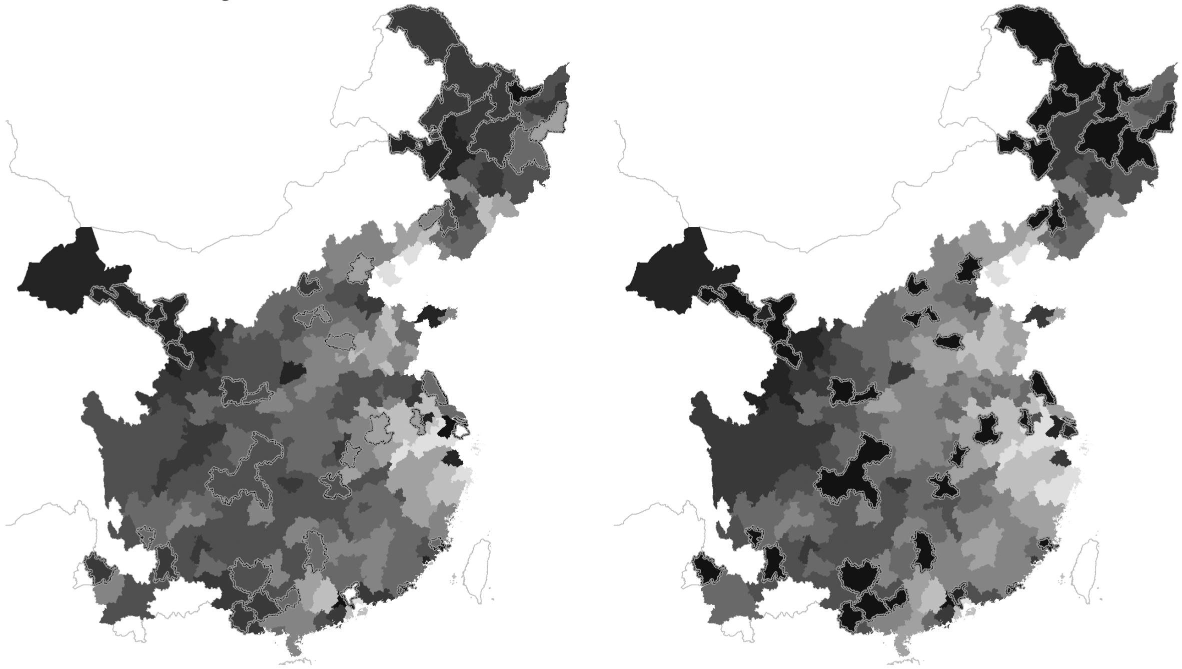
(d) Market Access

Figure 3: χ^2 test statistics of regional primate interaction coefficients in iv model jointly equal to zero.



Notes: Graph of test statistics comparing empirical models with primate city interactions to those without for different definitions of primacy. Primacy is defined as the highest population prefecture within the indicated number of minutes' drive over the 1962 road network at 90 kph. The blue line uses log 2010 GDP as the outcome and the red line uses log 2010 population as the outcome.

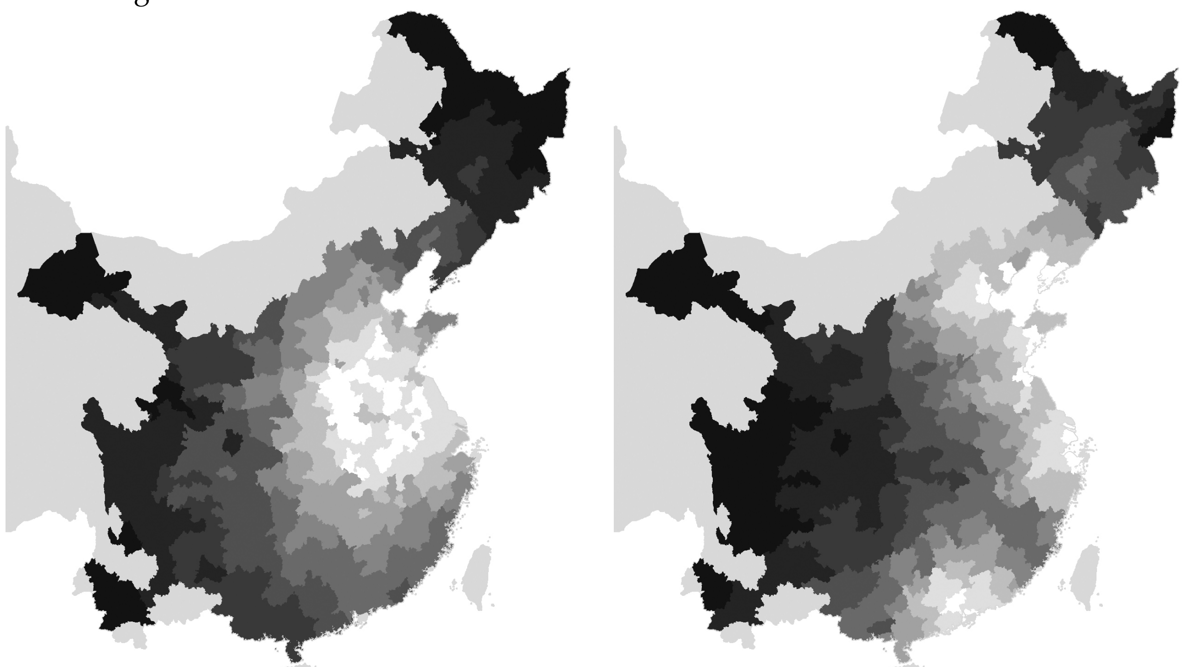
Figure 4: Geographic distributions of counterfactual population changes. In both panels lighter colors indicate larger values.



(a) Counterfactual changes in population no regional center effects

(b) Counterfactual changes in population with regional center effects

Figure A1: Geographic distributions of components of market access. In both panels lighter colors indicate larger values.



(a) Market Access Domestic Component (b) Market Access Trade Component

Table A1: Summary Statistics

Means and (Standard Deviations)

Log 2010 Road Efficiency	10.72
Units within 450 km	(0.40)
Log 2010 Time to Nearest	5.87
Port	(1.30)
Log 2010 Market Potential gravity	12.92
	(0.01)
Log Total Market Access	6.52
	(0.05)
Log Domestic Market Access	6.23
	(0.04)
Log External Market Access	5.13
	(0.06)
Log 1962 Roads Within	9.39
450 km, Excluding Own Prefecture	(0.29)
Log 1962 Time to Nearest	6.07
Port, Given Road Upgrades	(1.42)
Primate Prefecture Indicator	0.09
(Largest 1982 Pref Pop In a 300 Minute Drive)	(0.29)

Notes: Each statistic is calculated for 285 observations.

Table A2: OLS Infrastructure Regressions

	Log Prefecture GDP 2010	Log Prefecture Pop 2010	Prefecture Pop Gr. Rate 1990-2010	Log Prefecture GDP PerCap 2010	Log Private Firm Wage 2007
Log 2010 Road Efficiency	0.082	-0.077	-0.079*	0.16	-0.14**
Units within 450 km	(0.13)	(0.064)	(0.043)	(0.10)	(0.056)
Log 2010 Minimum Port	-0.14**	-0.098*	-0.067**	-0.039	-0.029*
Travel Time	(0.064)	(0.051)	(0.028)	(0.024)	(0.016)
Log Prefecture Area, 2005	-0.016	-0.047	-0.041	0.031	-0.075**
	(0.061)	(0.029)	(0.026)	(0.053)	(0.032)
Log Central City Area, 1990	-0.098*	-0.031	-0.023	-0.067*	0.023
	(0.051)	(0.027)	(0.017)	(0.039)	(0.020)
Log Central City Population,	0.12**	0.025	0.026	0.092*	0.018
1982	(0.055)	(0.026)	(0.018)	(0.047)	(0.029)
Log Central City Roughness	-0.051	0.0015	0.0022	-0.052*	0.036**
	(0.034)	(0.015)	(0.010)	(0.027)	(0.016)
Log Prefecture roughness	-0.019	0.0014	0.0043	-0.021	0.012
	(0.028)	(0.012)	(0.0095)	(0.022)	(0.015)
Provincial Capital	0.63***	0.35***	0.25***	0.28***	0.17**
	(0.11)	(0.053)	(0.038)	(0.089)	(0.071)
Log Prefecture Population,	0.54***	0.82***	-0.11***	-0.28***	0.054
1982	(0.087)	(0.050)	(0.031)	(0.065)	(0.041)
Share Prefecture Population,	0.68	-0.13	-0.27	0.81	-0.59
with High School, 1982	(0.95)	(0.44)	(0.35)	(0.72)	(0.56)
Share Prefecture Population,	2.16***	-0.32	0.035	2.48***	0.55**
in Manufacturing, 1982	(0.57)	(0.38)	(0.23)	(0.36)	(0.24)
Log km to Coast	-0.034	-0.012	-0.0086	-0.022	-0.014
	(0.035)	(0.013)	(0.011)	(0.029)	(0.013)
West Region	-0.041	0.012	0.0077	-0.054	-0.0079
	(0.11)	(0.042)	(0.034)	(0.090)	(0.053)
East Region	0.18**	-0.041	-0.024	0.22***	0.092**
	(0.084)	(0.046)	(0.031)	(0.066)	(0.044)
Constant	-1.85	4.79***	3.20***	-6.64***	10.8***
	(2.15)	(1.51)	(0.86)	(1.24)	(0.75)
R-squared	0.78	0.89	0.43	0.63	0.33

Notes: Regressions are analogous to those in Table 2.

Table A3: OLS Infrastructure Regressions With Primate City Interactions

	Log Prefecture GDP 2010	Log Prefecture Pop 2010	Prefecture Pop Gr. Rate 1990-2010	Log Prefecture GDP PerCap 2010	Log Private Firm Wage 2007
Log 2010 Road Efficiency	0.022	-0.11	-0.11**	0.13	-0.19***
Units within 450 km	(0.14)	(0.072)	(0.047)	(0.11)	(0.061)
X Rank 1 Prefecture	0.39**	0.26***	0.22***	0.13	0.25**
	(0.18)	(0.092)	(0.065)	(0.14)	(0.12)
Log 2010 Time to Nearest	-0.16**	-0.11*	-0.072**	-0.050*	-0.025
Port	(0.073)	(0.059)	(0.032)	(0.025)	(0.018)
X Rank 1 Prefecture	0.095	0.045	0.020	0.050	-0.036
	(0.075)	(0.046)	(0.026)	(0.050)	(0.024)
Rank 1	-4.72**	-2.98**	-2.41***	-1.75	-2.51*
	(2.15)	(1.17)	(0.77)	(1.60)	(1.33)
R-squared	0.78	0.89	0.45	0.63	0.35

Notes: Regressions are analogous to those in Table 4.

Table A4 Market Potential and Market Access Results by Prefecture Primacy

	Log Prefecture GDP 2010	Log Prefecture Pop 2010	Growth Prefecture pop 1990-2010	Log Prefecture GDP pc 2010	Log Private Firm Wage 2007
Panel A: Market Potential Gravity Regressions					
Log 2010 Market Potential	-6.90	-9.58**	-9.64***	2.67	-9.14**
Gravity	(8.39)	(4.32)	(3.32)	(6.46)	(3.95)
X Rank 1 Prefecture	22.7**	14.3***	12.5***	8.40	12.8**
	(10.7)	(4.78)	(3.77)	(8.53)	(6.42)
Log 2010 Time to Nearest	-0.18**	-0.11*	-0.074**	-0.068**	-0.033*
Port	(0.076)	(0.062)	(0.034)	(0.028)	(0.020)
X Rank 1 Prefecture	0.11	0.050	0.025	0.057	-0.039
	(0.087)	(0.052)	(0.031)	(0.057)	(0.034)
Rank 1 Prefecture	-294**	-185***	-162***	-109	-165**
	(138)	(61.9)	(48.8)	(110)	(83.0)
Panel B: Market Access Regressions					
Log 2010 Market Access,	-9.37**	-6.99**	-5.39***	-2.39	-3.76***
Domestic	(4.60)	(3.42)	(1.90)	(2.17)	(1.45)
X Rank 1 Prefecture	4.69	2.35	1.85	2.34	0.42
	(4.42)	(2.23)	(1.42)	(3.25)	(2.27)
Log 2010 Market Access,	13.6**	8.53*	5.77**	5.09**	2.97**
External	(5.86)	(4.72)	(2.51)	(2.11)	(1.49)
X Rank 1 Prefecture	-2.98	-0.97	-0.19	-2.01	2.59*
	(3.47)	(2.17)	(1.29)	(2.68)	(1.56)
Rank 1 Prefecture	-13.9	-9.58	-10.5**	-4.31	-16.0*
	(16.1)	(7.14)	(5.09)	(10.9)	(8.18)

Notes: Regressions are the same as those in Table 5 Panels A and C, except with the addition of primate city interactions. The first stage F-statistic is 26.2 in Panel A and 10.5 in Panel B.