

**Nathaniel Baum-Snow, [J. Vernon Henderson](#), Matthew A. Turner and Loren Brandt**

## **Does investment in national highways help or hurt hinterland city growth?**

**Article (Accepted version)  
(Refereed)**

**Original citation:**

Baum-Snow, Nathaniel and Henderson, J. Vernon and Turner, Matthew A. and Brandt, Loren (2018) *Does investment in national highways help or hurt hinterland city growth?* [Journal of Urban Economics](#). ISSN 0094-1190 (In Press)

© 2018 Elsevier B.V.

This version available at: <http://eprints.lse.ac.uk/88087/>

Available in LSE Research Online: May 2018

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

# Does Investment in National Highways Help or Hurt Hinterland City Growth? \*

Nathaniel Baum-Snow, University of Toronto

J. Vernon Henderson, London School of Economics

Matthew A. Turner, Brown University      Qinghua Zhang, Peking University

Loren Brandt, University of Toronto

April 26, 2018

## 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 local 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 the International Growth Centre for helping fund this research. We also acknowledge the support of the Global Research Program on Spatial Development of Cities at LSE and Oxford University funded by the Multi Donor Trust Fund on Sustainable Urbanization of the World Bank and supported by the UK Department for International Development. We received helpful comments from Gerald Carlino, Edward Glaeser, Samuel Marden, Daniel Sturm, Junfu Zhang, two anonymous referees, 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 network on China’s economic geography. In particular, we examine effects on the spatial distributions of population, GDP, wages, and the composition of output around year 2010. Our investigation focuses on how highways’ promotion of improved access to regional domestic markets affects these outcomes and how these effects depend on a prefecture’s location in the regional urban hierarchy. We also separately consider effects of highways that provide improved access to international markets.

Our investigation faces three 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 and predates reliance on roads to transport goods between prefectures.

Second, we must measure a prefecture’s location in the urban hierarchy. To measure a prefecture’s location in the urban hierarchy we define ‘regional primate prefectures’ as the highest population 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. We also examine pure distance based measures.

Third, we must measure the relationship between the road network and market access. This raises difficult problems for estimation. Theoretically founded definitions of market access are fundamentally recursive. If improved access to prefecture B from prefecture A increases the size of prefecture A’s economy, then the converse relationship should also hold. However, this implies that shocks to prefecture A’s economy affect prefecture A recursively through prefecture B. This raises obvious challenges for the estimation of the causal effects of market access on local economic outcomes. This is a natural implication of general equilibrium, and in theory, can be solved with the specification of a correct structural model of the economy. Our contribution is to identify important stylized facts about China’s economic geography that such a model should reflect, and to a lesser extent, to point out that these facts are not obviously consistent with several widely used models of economic geography.

We skirt this problem by primarily considering measures of market 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 the minimum travel time to a major international port along the highway network. Since these measures do not depend on economic activity, they avoid the recursion problem. They also measure quantities to which policy makers can directly relate. Our various measures of market access are sufficiently highly correlated that we cannot empirically distinguish between the treatment effects of these alternatives.

In a robustness check, we report on outcomes in which we use the traditional gravity measure of market access which we call 'market potential': the inverse of a travel time weighted sum of economic activity around a prefecture. 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, does introduce the recursion problem described above. We also report on results using a market access measure derived from a Ricardian model of the sort now common in the economic geography literature (Donaldson and Hornbeck, 2016; Tombe and Zhu, 2015). However, since our work is not founded in that model and we wish to remain agnostic about the underlying data generating process, this is not the focus of our analysis.

To estimate causal effects of 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. To avoid the recursion problem, for instruments we rely on the same quantity based 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 market potential and market access variables. However since, instruments do not vary with specification, we cannot distinguish statistically between our preferred road quantity measure of local access and a gravity measure.

Our investigation leads to one central set of findings. Improved access to domestic markets reduces prefecture population, GDP, population growth and wages paid by private sector firms on average, although GDP effects are not significant. However, these average effects mask differences in how roads affect prefectures at different positions in the regional hierarchy. The negative effects of better access to local markets apply only to non-primate cities. Regional primate prefectures exhibit positive offsetting effects for populations, GDP and wages with improved domestic market access. For example, a 10 percent increase in roads within 450 km of a prefecture city reduces non-primate prefecture population by 1.7 percent but increases primate prefecture population by 1.1 percent. As hinterland prefectures shrink with better access to domestic markets they become relatively more specialized in agriculture at the expense of manufacturing and services. These effects 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, whether the prefecture is a nodal point on the highway network, or a provincial capital. Finally we also look at the effect of better access to nine coastal ports. In general, better connections are associated with increased GDP and population for all cities regardless of position in the urban hierarchy.

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 relatively concentrate people in regional primates. With this migration comes an increase in output in the regional primates, manufacturing in particular, while hinterland prefectures shrink and specialize relatively more in agriculture.

As a large, developing market economy with newly constructed infrastructure, cities and regions in China exhibit the variation required to study the effects of roads on regional development. While much about China is unique, there seems to be no particular reason to think our results will not apply to other developing market-based economies. Extrapolating our results to other developing world countries suggests that expansions of the road network favor the rise of regional primate cities over smaller cities elsewhere in the developing world.

These results are important for several reasons. First, billions of dollars of transportation infrastructure are under construction or consideration in the developing world. 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 centers 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 national strategic 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 these migration responses may go along with overall welfare improvements, they are the opposite of what is intended.

Second, 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 (Krugman, 1991) and subsequent generalizations (Puga, 1999; Fujita, Krugman and Venables, 1999; see Ottaviano and Thisse, 2004, for a review), the literature has developed a two area model of economic geography with limited population mobility between the core (our primate city) and periphery (our hinterland cities) areas. A decline in trade costs may lead to an increase in core population at the expense of the hinterland. The idea in these new economic geography [NEG] models is that there is a home market effect which can be amplified some by population mobility. With high transport costs, producers in the periphery enjoy a degree of trade protection which is reduced when transport costs fall and periphery residents gain by importing certain products from the core, which they had bought before from local producers. The result is a shift of employment to the core, so its population and GDP rise. This modeling context applies well to China. Tombe and Zhu (2015) present evidence of low population mobility rates between regions in China and while there is greater between prefecture mobility within regions, migration remains very costly.

Existing empirical evidence on the effects of highways on economic geography in developing countries is mixed. For China, like us Faber (2014) concludes that rural (periphery)

Chinese prefectures are hurt by better proximity to new highways. However, Banerjee, Duflo and Qian (2012) conclude that proximity to a highway or railroad is *beneficial* for an average Chinese county. 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. We provide more direct econometric evidence for an ‘urban hierarchy’ at the regional level, and the contrast between core and periphery outcomes.<sup>1</sup>

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 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.

There is a large literature which takes a different approach to examining the effects of national transport systems (e.g. Donaldson, 2015; Donaldson and Hornbeck, 2016; Alder, 2016; Sotelo, 2015; Allen and Arkolakis, 2014; Bartelme, 2015; Fajgelbaum and Redding, 2014; Tombe and Zhu, 2015; Balboni, 2017), including an early version of this paper (Baum-Snow, Henderson, Turner, Zhang and Brandt, 2016).<sup>2</sup> After experimenting with calibrations of standard versions of these models, we concluded that our results were not consistent with these models, as we discuss below. 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 industrial specialization that depends on prefecture abundance of land or natural resources (or on an exogenous comparative advantage in manufacturing and service sectors). The recent literature generally assumes returns to scale are ‘small enough’ to rule out multiple equilibria,<sup>3</sup> while land or natural resource abundance typically plays a small role in most structural models applied to transportation.<sup>4</sup>

---

<sup>1</sup>Our findings help reconcile apparent contradictions in the literature investigating the effect of roads and highways in China. Our results suggest these differences are a consequence of sampling. Faber (2014) deliberately oversamples rural prefectures, while Banerjee, Duflo and Qian (2012) do not.

<sup>2</sup>There are also papers on the role of transportation infrastructure and trade costs in economic development. Topalova and Khandelwal (2011) provide evidence that lower trade costs have 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.

<sup>3</sup>For an example of such a condition, see Theorem 2 (iii) and Proposition 1 in Arkolakis and Allen (2014).

<sup>4</sup>Nagy (2017) is an exception.

Thus, the focus of this paper is to establish facts about the impacts of the expressway system on China’s economic geography. Our object is to identify what forces determine the winning and losing regions from the reshuffling of economic activity caused by an expressway system. This is critical to public policy decision making and informs the consideration of modeling strategies that may rationalize our pattern of results.

## 2 Context and Data

The Chinese context is well-suited for our investigation. First, China is large and geographically varied enough to permit the emergence of 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.

Figures 1a and 1b show the national road networks in 1962 and 2010. We use these two networks to calculate quantity based measures of road infrastructure, e.g., kilometers of roads within 450km of a prefecture center,<sup>5</sup> and to estimate the cost of travel between any pair of prefectures.<sup>6</sup> These pairwise cost estimates, which we discuss in detail in section 2.4, are based on estimates of network travel time along the road network travel calculated assuming speeds of 25 kph on local highways and 90 kph on expressways. The lightly shaded region in figure 1 is our study area. We use the 285 prefectures in this area as our primary estimation sample. 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. We postpone a detailed discussion of our estimation strategy and instrument validity to section 3.

### 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

---

<sup>5</sup>For the purpose measuring infrastructure, we include roads within China that are outside our study area.

<sup>6</sup>We calculate pairwise travel times using the ARCGIS network analyst, which is based on the Dijkstra algorithm.

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. For the period 2000 to 2005, Tombe and Zhu (2015) find extremely high inter-provincial costs of moving, but also very high costs even for within province moves. Of course there is movement, even if mostly local; and China's share of population which is urban has risen to about 50% in 2010 from about 30% in 1990.

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 models of economic geography typically predict the effects of trade costs on population, output and wages, these are our primary outcomes of interest. Specifically, log 2010 population and log 2010 GDP are our primary outcomes, with log 2007 private sector wages as a measure of output per worker. As a robustness check, we also consider 1990-2010 population growth rates.<sup>7</sup> 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. From NEG models based on Krugman (1991), as noted earlier, we expect better highway connections to increase primate city populations and GDP and to reduce that for hinterland cities. The impact of wages is more ambiguous and model specific. In Krugman (1991) with a perfectly mobile population, core city nominal wages fall relative to those in the periphery, since reductions in trade costs reduce the periphery's price index more. Introducing local housing costs which rise with population (Helpman, 1998) and with more explicit modeling of migration costs (Balboni 2017), effects on nominal wages can be positive in the core region.

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

---

<sup>7</sup> All of our population measures are based on census data and reflects counts of people in a prefecture rather counts of people with Hukou registration in a prefecture.



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 data 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 GDP 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. This is an intuitively reasonable travel time cutoff, 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 prefecture population of 2.8 million. We note that a number of the smaller primate prefectures are in remote places poorly served by roads. Unsurprisingly, the top four 1982 population prefectures are all primates.

As robustness checks we consider a related continuous measure of regional primacy and examine distance rather than driving time based measures to define primate cities. For the continuous 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. For distance based measures we perform the same type of structural break test to identify the critical distance for primacy and rerun specifications based on this

definition of regional primates. In both cases, we obtain qualitatively similar results to the main exercise.

## 2.4 Roads and Travel Time

To describe the Chinese road 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 each of the nine most important international ports, and select the shortest one.<sup>8</sup> 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, as noted, we can and do examine population changes from 1990 to 2010.

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.<sup>9</sup>

Our measures of 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  $\tau_{ij}$ , we use

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

This expression captures both the pecuniary and time (opportunity) cost of shipping and incorporates some concavity. All reported results are based on  $\rho = 1$ . However, 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  $\tau_{ij}$  based on values of  $\rho$  between 0.5 and 2.

---

<sup>8</sup>The nine ports that handle the largest volume of international trade in 2001 were: Tianjin, Qinhuangdao, Dalian, Shanghai, Lianyungang, Ningbo, Qingdao, Guangzhou, and Shenzhen.

<sup>9</sup>We use ESRI's network analyst for these calculations. This software relies on the widely used Dijkstra algorithm to find routes that minimize travel times.

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. For reference, when  $\rho = 1$ , our expression for  $\tau$  requires a loss of 2.1% of value for an eight hour travel day.

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. Specifically, to calculate  $\tau_{ix}$  we use

$$\tau_{ix} = 1.15\tau_{ip} \quad (1)$$

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. We weight limited access highways more heavily in our efficiency kilometers because bigger roads accommodate more people and freight; the chosen weights reflect a rough guess at speed of travel along the roads. Regional variation in this efficiency unit measure is depicted on a map 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 for policy purposes.<sup>10</sup>

---

<sup>10</sup>We note that our measure of efficiency km does not correct for land area. Thus, places near the coast may have relatively few roads despite dense networks. This may partly explain lower values of efficiency roads along the central coast in figure 2a. Presumably, the same issue is also relevant for other infrastructure measures as well.

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). In robustness checks we show results for a 300 km radius. In general, our results are robust to different choices of radius, provided they are small enough to preserve cross-prefecture variation in the measure of efficiency kilometers. However, we do not have sufficient statistical power to separately estimate effects of infrastructure by distance ring.

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  $\tau_{ix}$  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. To capture both the domestic and international market access components to transport improvements, the access to international market measure is paired either with local road efficiency units or with the measure of market potential discussed next in relevant specifications.

## Market Potential

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

As a robustness exercise, we construct the following traditional gravity measure of ‘market potential’, the discounted sum of GDP surrounding each prefecture.

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

Theoretical foundations for this sort of formulation of market potential include Redding and Venables’ (2004), Hanson’s (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  $\sigma$ .

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 spatial distribution of market potential.

Note that since our instruments are the same for the road efficiency unit and market potential measures, we cannot distinguish between them statistically. We prefer the former for its direct policy interpretation. With that caveat in mind, in the results section and

Table A4 we also report results from market access formulations based on either the Eaton-Kortum (2002) paradigm used in, for example, Hornbeck and Donaldson (2016) which assumes perfect population mobility across areas or the NEG paradigm (Redding, 2016) with limited population mobility (Balboni, 2017; Tombe and Zhu, 2015). The Appendix lays out the construction of our measures of market access. A heat map for market access looks very much like Figure 1c for market potential.

### 3 Econometric Framework

As we have discussed, Krugman (1991) type models suggest that causal effects of transportation cost changes on prefecture GDP and population differ by position in the urban hierarchy. Effects on wages are more ambiguous. We estimate causal relationships between these outcomes and efficiency km of roads and various alternative 'market access' measures.

#### 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.

$$Y_{it} = a + \beta L_{it} + \psi E_{it} + X_i \delta + u_{it} \quad (3)$$

$$L_{it} = a_1 + \beta_1 L_{i62} + \psi_1 E_{i62} + X_i \delta_1 + \eta_{it}^1 \quad (4)$$

$$E_{it} = a_2 + \beta_2 L_{i62} + \psi_2 E_{i62} + X_i \delta_2 + \eta_{it}^2 \quad (5)$$

It is possible that some elements of  $u_{it}$  are correlated with  $L_{it}$  and  $E_{it}$  in equation (3). For example, more productive prefectures may have more resources to build highways. But higher productivity also directly generates greater GDP, population and wages. Other mechanisms such as prefecture government competency may also be a source of important omitted variables.

Incorporating the equations (4) and (5) 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  $E[L_{i62}u_{it}] = 0$  and  $E[E_{i62}u_{it}] = 0$ , 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.<sup>11</sup> 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 of interest  $\beta$  (Kelejian and Prucha, 2010).<sup>12</sup> 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.

### 3.2 Instrument Validity and First Stages

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.

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 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.

The People’s Daily (June 11, 1963) describes a major road building effort undertaken in the early 1960’s; ‘The present effort at building roads aims at opening up commercial routes to the villages to facilitate the transport of locally-produced goods as part of the policy of priority given to agriculture. Better roads are being built by provincial governments, but most of them are being built at local initiative. They are rarely fit for motor traffic; on

---

<sup>11</sup>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 (3) gives  $\ln Y_{1t} = a + \beta \ln((a + \beta \ln(Y_1/\tau_{12}) + \psi E_{2t} + X_2\delta + u_{2t})/\tau_{12}) + \psi E_{1t} + X_1\delta + u_{1t}$ .

<sup>12</sup>Gibbons, Overman and Pattacchini (2015) discuss the pitfalls of using these methods. In particular any heterogeneity in  $\beta$  would render all parameter estimates recovered using a standard spatial lag estimator inconsistent.

the better roads horses and ox-carts may travel; on others hand-carts can be pushed or pulled by man' (Lippit, 1966 p. 115).

To sum up, the 1962 road network was not suitable for moving manufactured goods or people with late 20th century technology. Moreover, it was organized to fulfill the historical objective of moving agricultural goods from the countryside into the nearby urban center, and not to facilitate the movements of goods and people that characterize a modern industrial economy. In spite of this, historical roads provide rights-of-way facilitating lower cost highway construction over or alongside old roads, all of which has taken place since 1990. Therefore, we expect the 1962 road network to indicate routes where roads can be constructed at low cost, but that it will not reflect late 20th century travel and shipping demand.

We use the 1962 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.<sup>13</sup> The rationale noted above for these instruments is based on the idea that 1962 roads were built for other reasons but, but even low quality ones were upgradeable to modern highways at lower cost than would be required to establish new rights of way. As a result of this lower cost, *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 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.<sup>14</sup> Because 1962 roads primarily served as connections from agricultural areas to nearby cities, we also control for urbanization with 1982 prefecture city population. We

---

<sup>13</sup>Changes to the speed of travel along the 1962 network rescale the regression coefficients. Choosing the same speed for 1962 makes it easier to compare travel times across years.

<sup>14</sup>1982 is the first year for which we have information on these variables. Using the 1982 census as controls in our regressions raises the possibility that the 1962 road network affects subsequent outcomes through these controls. For this reason, if 1962 demographics were available, we would prefer them. With this said, China was relatively static during the period 1962-82. The earliest market reforms (which affected only agriculture) did not occur until the early 1980s, and economic growth was about half as fast during the period 1962-82 as during 1982-2002. Thus, our implicit claim that 1962 roads affect post-1990 development only through their effect on road construction post-1990, and not through their effect on 1982 demographics seems plausible; China was not changing very rapidly during this period so 1962 roads had limited scope to affect 1982 demographics.

control for prefecture and central city roughness to proxy for agricultural productivity. To estimate effects of shorter 2010 driving time to the nearest of 9 coastal ports, we instrument with driving times over 1962 roads to the respective coastal port. This requires a control for distance to the nearest point on the coast. Distance to the coast controls for many things including access to Eastern non-port cities which are politically important like Beijing. We note however that if, instead of distance to the coast, we control for distance to the nearest of these 9 major ports, that removes any power of this instrument. For this reason, we primarily focus our discussion on effects of domestic road linkages. Finally, controlling for provincial capital status accounts for the fact that these cities have distinct institutional and industrial histories.

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. The result in column 3 for the alternative measure of local market access, market potential, is also strong, even though the instrument does not contain information on GDP of other prefectures.

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 the dependent variable but assuming 90 kph travel speeds over the 1962 network. 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.

Columns 1 to 3 of Table 1 show that our instruments are strong. These results also confirm our expectation that the 1962 regional roads instrument predicts 2010 efficiency km, while 1962 travel time to port predicts its modern counterpart.



## 4 Results

### 4.1 Average effects of efficiency km of roads within 450 km and travel time to port

Table 2 reports coefficient estimates based on the regression equation (3), 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 have four prefecture outcomes: 2010 GDP, 2010 population, 1990 - 2010 population growth and 2007 wages of private sector industrial firms.<sup>15 16</sup>

We first consider effects of travel time to a major international port, seen in the second row of Table 2. As expected, the average effect of 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. We also find a positive effect on private sector wages, with this elasticity estimated at -0.04.

Estimated effects of regional road capacity, in the first row of Table 2, are perhaps surprising because they are all negative. While the effect on GDP is not significant, the others are. In columns 2 and 3, 10% more road capacity nearby leads to about 1.2% smaller prefecture population, or has a strong dispersion effect. Without controls, results not shown indicate that the relationships between regional roads and both population and GDP are positive, but for growth (where fixed historical conditions might be viewed as being differenced out), the coefficient remains negative and significant and little changed at -0.11. That is, 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. Column 4 shows a negative effect on wages. Table A2 reports analogous OLS regressions. OLS results are qualitatively similar to IV ones, but with smaller dispersion effects.

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

---

<sup>15</sup>We note that the interpretation of regressions of *change in population* on roads are quite different from the interpretation of a regression of *population* on roads. In fact, since the modern road network was constructed entirely since 1990, variables describing the level of roads in 2010 and also describe changes since 1990. Thus, formally, by changing the dependent variable alone, we convert the regression from a levels regression to a first difference regression. It follows that we can interpret coefficients in the two regressions in the same way.

<sup>16</sup>In theory our data permit us to do first difference regressions between 2010 and 2000 as well. While first difference estimation has well known advantages, in practice, this is not possible for us. Our instruments do not predict the changes in road infrastructure from 2000 to 2010, particularly the travel time to the nearest port.

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. Results in remaining regression tables have the same controls as in Table 2. We do not report their coefficients as they follow similar patterns.

The focus of the rest of the paper is to explore these negative average treatment effects of improved roads. We note that if, in Table 2, we replace log 2010 road efficiency units with the measure of log market potential, the coefficients (and standard errors) in that row for Columns 1-4 on GDP, population, population growth and wages are respectively -1.88 (7.76), -7.25\*\* (3.57), -7.54\*\*\* (2.83) and -6.66\* (3.62). Similarly we can take a measure of market access from either the Eaton-Kortum or NEG structural literature, decomposed into domestic and international components as given in the Appendix. While improved international market access has positive effects in Table A4 Panel B, improved domestic market access has negative effects which are significant at the 5% level in all columns except GDP and there it is at the 10% level. These of course contradict predictions from these models. The key take away is that our road efficiency unit measure is not driving these negative average treatment effects of improving local market access.

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

We now show that effects of improved regional road infrastructure are related to a 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 3a shows  $\chi^2$  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  $\chi^2$  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  $\chi^2$  statistic does not vary with driving radius and is everywhere below levels that indicate interactions are statistically significant. 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. Below we will also discuss results for a distance based measure of primate cities.

We note the obvious similarity between our method of defining primate cities and structural break tests. As the radius increases, the risk of missing the actual primate decreases, yet it becomes more likely that one mis-identifies a large prefecture that is too distant to exert much influence on the hinterland prefecture. One balances these two risks at the structural break.<sup>17</sup>

Table 3 Panel A reports regressions analogous to those in Table 2, but with infrastructure variables interacted with urban primacy. The negative effects of efficiency km of roads on population, population growth and wages seen in Table 2 are about 50% greater in magnitude for non-primate cities, with the negative effects for GDP larger also. In contrast, regional primates experience statistically significant offsetting positive effects for all outcomes. The first take-away is that relative effects are very different between the two sets of cities. Second, the sum of primate and non-primate coefficients is positive and reasonably large in all cases. The net marginal effects for regional primates appear positive, although the sum is generally only weakly significant.<sup>18</sup> 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.

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 3 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 examined. Interaction terms for the continuous primacy variable are positive and highly significant for GDP and population.

Table 3 indicates that regional primates are affected differently by regional roads than are hinterland prefectures. However, evidence on differential effects of port connections is not present in 3 of the 4 cases in Panel A, though Panel B exhibits some weakly significant differentials. Taken together, this suggests that regional primates are less affected by the cost of trucking goods to international markets than are hinterland prefectures. Section 4.4 provides some corroborating evidence.

Results in Table 3 are based on our preferred measures of transport from Table 2. We offer three robustness checks in Table 4. First in Panel A, we present corresponding results using market potential as the measure of local access. Results on differential effects for regional primates are similar to those in Table 3, except of course coefficients are rescaled to reflect the differences in units of measure. In particular, since the standard deviation

---

<sup>17</sup>We are grateful to a referee for suggesting this intuition.

<sup>18</sup>For example for GDP and population, Chi-sq statistics are 2.95 and 2.12 which for a one tail test have p-values of 0.086 and 0.145 respectively.

of market potential is much smaller than that of efficiency roads, coefficients for market potential measures tend to be larger. Then Panel B returns to road efficiency units within 450 km, but uses a primacy measure based on being the primate city within a certain distance (rather than driving time over 1962 roads). As shown in Figure 3b, the structural break here for both GDP and population appears to be at a distance of about 420 km, with just 9 regional primates emerging. The results in Panel B are very strong for GDP and population. For hinterland cities, more efficiency units result in losses in all cases, which are all significant except for GDP. Differentials for this more restricted set of primates for GDP, population and population growth are now very large with significant positive net marginal effects for these three outcomes. In Panel C we limit the road efficiency unit measure to roads within 300km rather than 450km. Results are qualitatively similar to those in Table 3.

### 4.3 Primacy versus absolute size, political status, and transport node

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 3 but with a different primacy definition in each panel.

In Panel A, we look at nodal cities in 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 decile population prefecture 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.

#### 4.4 Sector-Specific Effects Through the Hierarchy

We expect sectoral differences in responses to a better regional road network. For example, hinterland producers of traded goods with low land shares, high fixed costs, or that benefit more from agglomeration economies are arguably more likely to 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, as in the Krugman (1991) model. Agriculture has a high land share and so seems likely to respond in the opposite way. That is, hinterland areas should become more specialized in agriculture with a better regional road network. Non-traded services may 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 3, 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 3 Columns 2 and 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 contrast with 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 ac-

cess 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 employments in regional primates do 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.

## 4.5 Openness to trade and rail

Our study period also saw a dramatic increase in the extent to which China was open to trade with the rest of the world. Given this, we are concerned that increased openness to trade may partly explain our results. To address this issue, we assemble data identifying all prefecture cities that contained Special Economic Zones (SEZs) in 1995. These zones are designated parts of cities which enjoy a relaxed regulatory environment and, sometimes, favorable tax treatment. They are intended to attract foreign direct investment and to stimulate exports.<sup>19</sup> In short, prefectural cities containing and SEZ are particularly open to trade .

In order to assess the role of China's increasing openness to trade on our findings, Appendix Table A5 replicates the results of Table 3, panel A, while adding an indicator for SEZ status and the interaction of this indicator with the indicator for regional primate status. The SEZ indicator is highly significant and has the expected positive effect on output and population, while the interaction of SEZ status and primate status is not distinguishable from zero. Comparing with panel A of table 3, we see that including these two controls for SEZ status does not qualitatively affect our main results. Indeed, most coefficients in table 3 panel A change minimally and all by well under a standard error with the addition of the SEZ variables. This suggests that our main results are not driven primarily by changes in China's openness to trade.

While we have not discussed it, rail is also an important component of China's transportation network. While we have the same overtime detailed coverage of rails, the rail data do not account well for changes in rails such as the extension to double tracking, and perhaps the de facto closing of lines.<sup>20</sup> We have experimented widely with specifications that include measures of the rail network. These investigations have largely failed to reveal robust patterns in the data. We believe that this reflects the fact that regression specifications like our main specification (3)-(5) are simply too complicated when augmented with equations for rail, and that measures of the rail network and the road network are often highly correlated. This leaves open the possibility that some results may confound

---

<sup>19</sup>Our data on SEZs comes from the official website of China's Association of Development Zones, ([www.cadz.org.cn](http://www.cadz.org.cn))

<sup>20</sup>Our rail data are described in a companion paper, Brandt et al. (2017)

the effects of roads and railroads, though resolving this issue seems to be beyond the reach of our current data and research design.

## 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 7 shows the results. Column 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 the 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.

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 4 provides visualizations of these population results with and without differential effects for primate cities.

If we think of the counterfactual as offering up the effects from building the expressway system, *ceteris paribus*, we can compare the numbers with true population changes, 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. The counterfactual exercise has much greater magnitudes: a gain of 1.2 million for primates from building the system, with non-primates losing on average 121,000, *ceteris paribus*. This difference suggests that building expressways was just one of many changes in this dynamic 20 years for China, and other factors offset some of the relative gains for regional primates from expressway development.

## 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. 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. 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. These prefectures also become relatively more specialized in agriculture at the expense of manufacturing and traded 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.

Our conclusions are limited in one important way. Our reduced form methodology identifies the way that highways 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. 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 noted in the text and highlighted in appendices 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, opposite signs of domestic and international components of market access (reported in Appendix Table A4) seem hard to reconcile with the underlying theory, as does the heterogeneity in how access to markets affects 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 advantages. 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.

## Appendix A. Structural Market Access in Reduced Form Regressions

Changing trade costs between any two cities may affect trade flows between other pairs of cities. Neither efficiency km of roads within 450 km 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 can 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 (2017) all deliver such a measure. Full derivation of the variants of market access we use here is in Baum-Snow et al (2016). This derivation is done in the context of a standard Ricardian model (Eaton and Kortum, 2002) that follows closely from Hornbeck and Donaldson (2016); we also 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  $\tau_{ij}$  is pairwise transport cost as defined above. Finally,  $\theta$  is the dispersion parameter from Frechet distributed productivity draws, which determines the gains from trade between prefectures, with larger values of  $\theta$  indicating smaller gains from trade.

For the purposes of the reduced form empirics in this paper, the following expression for each city’s ”market access”  $MA_i$  is the key one:

$$MA_i = \sum_j \tau_{ij}^{-\theta} \frac{Y_j}{MA_j} + \frac{\tau_{ix}^{-\theta} E}{\sum_j \frac{Y_j}{MA_j} \tau_{jx}^{-\theta}}, i = 1, \dots, 285, \quad (6)$$

where

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

is the value of exports.

Equation (6) defines market access with a recursive equation, and given data on  $Y_i$ ,  $\tau_{ij}$  and  $E$  and calibration of  $\theta$ , we can solve this system of equations for  $MA$ . We refer to the first term in equation (6) 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 (6) 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 (6) numerically using the observed value of Chinese exports  $E$ , GDP in 2010  $Y_j$ , pairwise transportation costs  $\tau_{ij}$ , and set  $\theta = 5$  to obtain 285 values of  $MA_i$  and  $MA_x$ .<sup>21</sup>

In a Ricardian framework with perfect population mobility, increases in total market access increase both GDP and population, so signs on market access in regression equations as in Table A4, Panel A for population and GDP should be positive. In Table A4 Panel B, 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.<sup>22</sup> Evidence in Table A4 Panel B contravenes this prediction. Domestic market access effects are zero to negative whereas external market access effects are universally positive.

In China, the hukou system is known to constrain mobility. In Baum-Snow et al (2016), following Redding (2016) and Balboni (2017), 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  $\theta$  replaced by  $1 - \sigma$ , where  $\sigma$  is the elasticity of substitution in consumption, imply the same expression for market access as in equation (6). They also imply a positive constant elasticity relationship between wages (rather than population or GDP) and market access. Again the results on wages in Table A4 do not yield the predicted results on market access.

---

<sup>21</sup>We experimented with values of  $\theta$  ranging from 3 to 10. None of the results we report is sensitive to variation of  $\theta$  in this range.

<sup>22</sup>From equation (6), 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,  $\frac{d \ln(A+B)}{dx} = \frac{A}{A+B} \frac{d \ln A}{dx} + \frac{B}{A+B} \frac{d \ln B}{dx}$ .

## 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.
- Balboni, C. (2017) ‘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.
- Baum-Snow, N. V. Henderson, M.A. Turner, Q. Zhang and L. Brandt (2016) ‘Highways, Market Access, and Urban growth in China,’ LSE, SERCDP0200.
- Baum-Snow, N. V. Henderson, M.A. Turner, Q. Zhang and L. Brandt (2017) ‘Roads, Railroads, and Decentralization of Chinese Cities,’ *Review of Economics and Statistics*, 99(3):435-448
- Bird, J. and S. Straub (2015) ‘The Brasilia Experiment: Road Access and the Spatial Pattern of Long-Term Local Development in Brazil,’ manuscript.
- Buera, F. and E. Oberfeld (2014) ‘The Global Diffusion of Ideas,’ manuscript.
- 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.
- 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.

- 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.
- 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.
- 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*.
- Helpman, E (1998) ”The size of regions,” in D. Pines, E. Sadka, I. Zilcha (Eds.), *Topics in Public Economics. Theoretical and Empirical Analysis*, Cambridge University Press (1998), pp. 33-54
- 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.
- Limao, N. and A. J. Venables (2001) ‘Infrastructure, Geographical Disadvantage, Transport Costs, and Trade,’ *World Bank Economic Review*, 15(3): 451-479.
- 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): 565-586.
- Nagy, D. K. (2017) ‘City Location and Economic Development,’ manuscript.
- Ottaviano and J. Thisse (2004) ‘Agglomeration and Economic Geography,’ *Handbook of Regional and Urban Economics*, 4: 2563-2608.
- Puga, D. (1999) ‘The rise and fall of regional inequalities,’ *European Economic Review*, 43(2): 303-334.
- Redding, S. and A. J. Venables (2004) ‘Economic Geography and International Inequality,’ *Journal of International Economics*, 62: 53-82.
- Redding, S. (2016) ‘Goods Trade, Factor Mobility and Welfare,’ *Journal of International Economics*, 101: 148-167.
- 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.

**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)
<b>Instruments</b>			
Log 1962 Roads Within 450km, Excluding Own Pref	1.05*** (0.038)	-0.26** (0.13)	0.018*** (0.0017)
Log 1962 Minimum Port Travel Time Given Road Upgrades	-0.024*** (0.0080)	0.76*** (0.061)	-0.00036 (0.00030)
<b>Controls</b>			
Log Prefecture Area, 2005	-0.052*** (0.019)	-0.053 (0.054)	-0.0029*** (0.00078)
Log Central City Area, 1990	0.0055 (0.012)	0.031 (0.051)	-0.000089 (0.00048)
Log Central City Population, 1982	-0.026* (0.015)	-0.0076 (0.071)	-0.0012** (0.00058)
Log Central City Roughness	-0.0060 (0.0097)	0.041 (0.050)	0.00011 (0.00036)
Log Prefecture roughness	-0.019** (0.0093)	-0.040 (0.036)	-0.00044 (0.00031)
Provincial Capital	0.066* (0.038)	0.048 (0.12)	0.0013 (0.0013)
Log Prefecture Population, 1982	0.071*** (0.023)	0.017 (0.081)	0.0033*** (0.00087)
Share Prefecture Population with High School, 1982	-0.78** (0.32)	-1.27 (0.98)	-0.013 (0.010)
Share Prefecture Population in Manufacturing, 1982	-0.25 (0.16)	-0.52 (0.58)	0.0016 (0.0047)
Log km to Coast	0.0030 (0.0068)	0.055* (0.028)	-0.00053** (0.00026)
West Region	-0.25*** (0.031)	0.054 (0.087)	-0.0045*** (0.0011)
East Region	-0.014 (0.023)	-0.17 (0.11)	0.00091 (0.00080)
Constant	1.03*** (0.37)	3.82** (1.53)	12.8*** (0.017)
R-squared	0.90	0.88	0.74

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

Notes: Regressions in columns 1-3 have 285 observations and that in column 4 has 283 observations. First stage regressions are in Table 1. Kleibergen-Paap first stage F statistics are 236 in 1-3 and 237 in 4. Robust standard errors are in parentheses.



**Table 3: Infrastructure Effects with Primate Prefecture Interactions**

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

Notes: Each regression has the same set of control variables as in Table 2. Regressions in columns 1-3 have 285 observations and that in column 4 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 to those for population. Robust standard errors are in parentheses.

**Table 4. Robustness Checks**

	Log Pref GDP 2010 (1)	Log Pref Pop 2010 (2)	Growth Pref Pop 1990-2010 (3)	Log Wage 2007 (4)
<b>Panel A: Market Potential Gravity Regressions</b>				
Log 2010 Market Potential Gravity	-6.90 (8.39)	-9.58** (4.32)	-9.64*** (3.32)	-9.14** (3.95)
X Rank 1 Prefecture	22.7** (10.7)	14.3*** (4.78)	12.5*** (3.77)	12.8** (6.42)
Log 2010 Time to Nearest Port	-0.18** (0.076)	-0.11* (0.062)	-0.074** (0.034)	-0.033* (0.020)
X Rank 1 Prefecture	0.11 (0.087)	0.050 (0.052)	0.025 (0.031)	-0.039 (0.034)
Rank 1 Prefecture	-294** (138)	-185*** (61.9)	-162*** (48.8)	-165** (83.0)
<b>Panel B. Distance Based Measure of Primacy</b>				
Log 2010 Road Efficiency Units within 450 km	-0.073 (0.13)	-0.13** (0.061)	-0.13*** (0.046)	-0.11* (0.062)
X Primate Prefecture	1.12*** (0.29)	0.60*** (0.17)	0.47*** (0.14)	0.27 (0.21)
Log 2010 Minimum Port Travel Time	-0.17** (0.067)	-0.098* (0.055)	-0.066** (0.029)	-0.038** (0.019)
X Primate Prefecture	0.18*** (0.055)	0.034 (0.039)	0.014 (0.023)	-0.032 (0.027)
Primate Prefecture	-12.9*** (3.07)	-6.49*** (1.80)	-5.14*** (1.53)	-2.81 (2.25)
<b>Panel C. Road Efficiency Units within 300 km</b>				
Log 2010 Road Efficiency Units within 300 km	-0.33* (0.20)	-0.27*** (0.091)	-0.26*** (0.073)	-0.20** (0.094)
X Primate Prefecture	0.52*** (0.19)	0.31*** (0.098)	0.26*** (0.082)	0.25** (0.11)
Log 2010 Minimum Port Travel Time	-0.19** (0.076)	-0.12* (0.061)	-0.079** (0.033)	-0.034* (0.020)
X Primate Prefecture	0.090 (0.078)	0.038 (0.049)	0.014 (0.029)	-0.052* (0.031)
Primate Prefecture	-5.69*** (2.14)	-3.21*** (1.17)	-2.59*** (0.90)	-2.30* (1.21)

Regressions in columns 1-3 have 285 observations and that in column 4 has 283 observations.  
Robust standard errors are in parentheses.

**Table 5: Effects Using Alternative Definitions of Primacy**

	Log Pref GDP 2010 (1)	Log Pref Pop 2010 (2)	Pref Pop Growth Rate 1990-2010 (3)	Log Private Firm Wage 2007 (4)
<b>Panel A: Primacy as Nodal Prefecture in the 5-7 Road Plan</b>				
Log 2010 Road Efficiency	-0.059	-0.12**	-0.14***	-0.14**
Units within 450 km	(0.12)	(0.055)	(0.045)	(0.061)
X Nodal Prefecture	-0.0013	0.0086	-0.011	-0.019
	(0.047)	(0.031)	(0.021)	(0.015)
Log 2010 Minimum Port Travel Time	-0.19**	-0.11**	-0.097**	-0.069**
	(0.085)	(0.048)	(0.039)	(0.029)
X Nodal Prefecture	0.10	0.033	0.055	0.041
	(0.090)	(0.055)	(0.040)	(0.029)
<b>Panel B: Primacy as Top 10% of 1982 Center City Populations</b>				
Log 2010 Road Efficiency	-0.085	-0.15**	-0.15***	-0.15**
Units within 450 km	(0.13)	(0.070)	(0.048)	(0.061)
X Large Population City	-0.047	-0.049	-0.026	-0.019
	(0.057)	(0.044)	(0.025)	(0.016)
Log 2010 Minimum Port Travel Time	-0.22*	-0.16*	-0.10*	-0.066**
	(0.12)	(0.096)	(0.053)	(0.029)
X Large Population City	0.15	0.11	0.068	0.055*
	(0.10)	(0.080)	(0.045)	(0.030)
<b>Panel C: Primacy as Provincial Capitals</b>				
Log 2010 Road Efficiency	-0.080	-0.14**	-0.14***	-0.12*
Units within 450 km	(0.13)	(0.065)	(0.048)	(0.063)
X Provincial Capital	0.25	0.11	0.12*	0.098
	(0.19)	(0.079)	(0.069)	(0.13)
Log 2010 Minimum Port Travel Time	-0.20**	-0.11*	-0.077**	-0.040*
	(0.081)	(0.067)	(0.036)	(0.023)
X Provincial Capital	0.15**	0.051	0.028	-0.0057
	(0.070)	(0.049)	(0.028)	(0.022)
<b>Panel D: Primacy as non Rank 1 Provincial Capitals</b>				
Log 2010 Road Efficiency	-0.022	-0.11*	-0.12***	-0.13**
Units within 450 km	(0.13)	(0.058)	(0.045)	(0.062)
X Non Rank 1 Provincial Capital	0.056	0.087	0.092	0.074
	(0.23)	(0.069)	(0.066)	(0.14)
Log 2010 Minimum Port Travel Time	-0.17**	-0.11*	-0.076**	-0.039**
	(0.070)	(0.057)	(0.031)	(0.020)
X Non Rank 1 Provincial Capital	0.15***	0.077**	0.056***	0.016
	(0.053)	(0.034)	(0.021)	(0.022)

Notes: First stage F-statistics are 8.3 in Panel A, 22 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 a 360 minute drive yield first stage F-statistics that are too small for resulting regression results to be informative.

**Table 6: Effects on log Employment by Industry**

	Total (1)	Agric. (2)	Manuf. (3)	FIRE & Bus. Svc. (4)	Other Services (5)
<b>Panel A: Binary Measure of Regional Prefecture Primacy</b>					
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)
<b>Panel B: Continuous Measure of Regional Prefecture Primacy</b>					
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 observations. The First stage F-statistic is 157 for each regression in Panel A and 285 for each regression in Panel B.

**Table 7: Impacts of Downgrading Expressways on Population  
Population Gains Associated with Going from 2010 to 1990 Roads Infrastructure**

	Highways become 25 kph (1)	Port travel time at 25 kph (2)	Both (3)
Full Sample	531,676 (624,731)	-538,052 (396,773)	0 (570,695)
Primate Prefectures Only	-748,171 (694,986)	-640,686 (572,957)	-1,206,162 (947,008)
Non-Primate Prefectures Only	660,154 (448,526)	-527,749 (374,587)	121,082 (332,091)

Notes: 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. Estimates are based on the regression specification in Table 3, Column 2. Each cell of the table reports prefecture means with standeard deviations in parentheses.

**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 360 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 Private Firm Wage 2007
Log 2010 Road Efficiency	0.082	-0.077	-0.079*	-0.14**
Units within 450 km	(0.13)	(0.064)	(0.043)	(0.056)
Log 2010 Minimum Port	-0.14**	-0.098*	-0.067**	-0.029*
Travel Time	(0.064)	(0.051)	(0.028)	(0.016)
Log Prefecture Area, 2005	-0.016	-0.047	-0.041	-0.075**
	(0.061)	(0.029)	(0.026)	(0.032)
Log Central City Area, 1990	-0.098*	-0.031	-0.023	0.023
	(0.051)	(0.027)	(0.017)	(0.020)
Log Central City Population, 1982	0.12**	0.025	0.026	0.018
	(0.055)	(0.026)	(0.018)	(0.029)
Log Central City Roughness	-0.051	0.0015	0.0022	0.036**
	(0.034)	(0.015)	(0.010)	(0.016)
Log Prefecture roughness	-0.019	0.0014	0.0043	0.012
	(0.028)	(0.012)	(0.0095)	(0.015)
Provincial Capital	0.63***	0.35***	0.25***	0.17**
	(0.11)	(0.053)	(0.038)	(0.071)
Log Prefecture Population, 1982	0.54***	0.82***	-0.11***	0.054
	(0.087)	(0.050)	(0.031)	(0.041)
Share Prefecture Population, with High School, 1982	0.68	-0.13	-0.27	-0.59
	(0.95)	(0.44)	(0.35)	(0.56)
Share Prefecture Population, in Manufacturing, 1982	2.16***	-0.32	0.035	0.55**
	(0.57)	(0.38)	(0.23)	(0.24)
Log km to Coast	-0.034	-0.012	-0.0086	-0.014
	(0.035)	(0.013)	(0.011)	(0.013)
West Region	-0.041	0.012	0.0077	-0.0079
	(0.11)	(0.042)	(0.034)	(0.053)
East Region	0.18**	-0.041	-0.024	0.092**
	(0.084)	(0.046)	(0.031)	(0.044)
Constant	-1.85	4.79***	3.20***	10.8***
	(2.15)	(1.51)	(0.86)	(0.75)
R-squared	0.78	0.89	0.43	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 Private Firm Wage 2007
Log 2010 Road Efficiency	0.022	-0.11	-0.11**	-0.19***
Units within 450 km	(0.14)	(0.072)	(0.047)	(0.061)
X Rank 1 Prefecture	0.39**	0.26***	0.22***	0.25**
	(0.18)	(0.092)	(0.065)	(0.12)
Log 2010 Time to Nearest Port	-0.16**	-0.11*	-0.072**	-0.025
	(0.073)	(0.059)	(0.032)	(0.018)
X Rank 1 Prefecture	0.095	0.045	0.020	-0.036
	(0.075)	(0.046)	(0.026)	(0.024)
Rank 1	-4.72**	-2.98**	-2.41***	-2.51*
	(2.15)	(1.17)	(0.77)	(1.33)
R-squared	0.78	0.89	0.45	0.35

Notes: Regressions are analogous to those in Table 3 Panel A.



**Table A4: IV Effects of Market Potential and Market Access**

	Log Prefecture GDP 2010	Log Prefecture Pop 2010	Prefecture Pop Gr. Rate 1990-2010	Log Private Firm Wage 2007
<b>Panel A: Market Access Regressions</b>				
log 2010 Market access	2.88* (1.60)	0.53 (0.92)	-0.17 (0.61)	-0.52 (0.70)
First stage F	69.2	69.2	69.2	68.8
<b>Panel B: Market Access Regressions, Domestic and External</b>				
log 2010 Domestic Market Access	-8.58* (4.48)	-6.63** (3.34)	-5.08*** (1.84)	-3.56** (1.42)
log 2010 External Market Access	13.0** (5.57)	8.15* (4.49)	5.61** (2.41)	3.47** (1.49)
First stage F	21.3	21.3	21.3	21.5

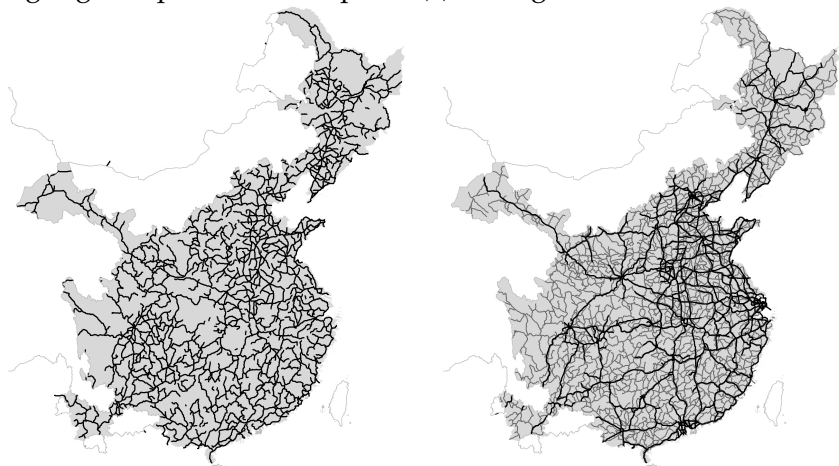
Notes: Each regression has the same set of control variables as in Table 2. Regressions in columns 1-3 have 285 observations and that in column 4 has 283 observations. First stage regressions use same instruments as in Table 3. Market access measures are from eqn. (6) of the Appendix. Robust standard errors are in parentheses.

**Table A5: Infrastructure Effects with Primate Prefecture Interactions and SEZ Controls**

	Log Prefecture GDP 2010 (1)	Log Prefecture Pop 2010 (2)	Prefecture Pop Gr. Rate 1990-2010 (3)	Log Private Firm Wage 2007 (4)
Log 2010 Road Efficiency Units within 450 km	-0.049 (0.14)	-0.14** (0.069)	-0.15*** (0.050)	-0.15** (0.066)
X Primate Prefecture	0.50** (0.23)	0.29*** (0.093)	0.25*** (0.065)	0.24** (0.12)
Log 2010 Minimum Port Travel Time	-0.14* (0.074)	-0.097 (0.061)	-0.063* (0.033)	-0.027 (0.020)
X Primate Prefecture	0.018 (0.080)	0.0095 (0.051)	-0.010 (0.029)	-0.067** (0.033)
Primate Prefecture	-5.36* (2.77)	-3.08** (1.22)	-2.57*** (0.79)	-2.23 (1.38)
1995 Special Economic Zone	0.40*** (0.089)	0.15*** (0.043)	0.12*** (0.037)	0.053 (0.041)
X Primate Prefecture	-0.054 (0.17)	-0.035 (0.074)	-0.038 (0.057)	-0.052 (0.11)

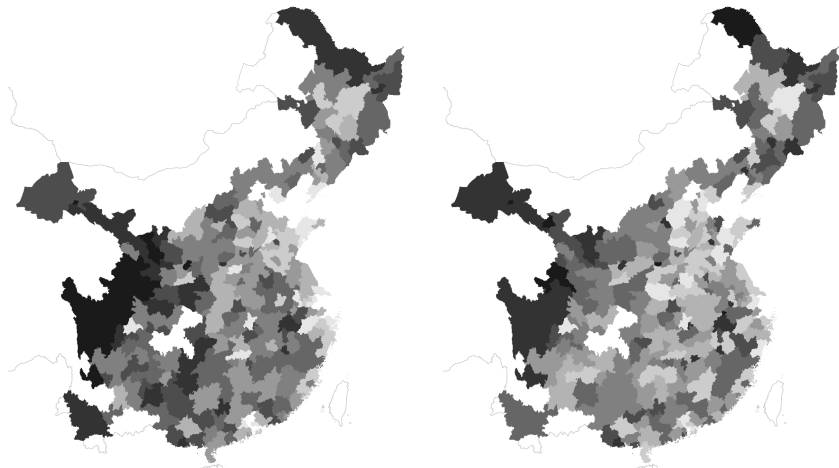
Notes: Each regression has the same set of control variables as in Table 2. Regressions in columns 1-3 have 285 observations and that in column 4 has 283 observations. The Kleibergen-Paap first stage F statistic is 143 for each regression. Robust standard errors are in parentheses.

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



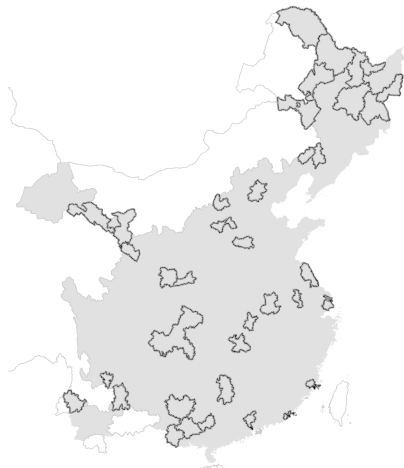
(a) 1962 road network

(b) 2010 road network



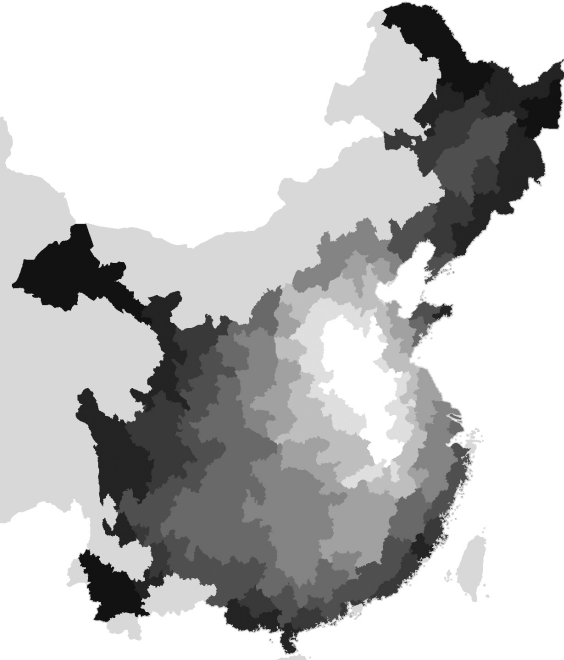
(c) 2010 GDP

(d) 2010 population

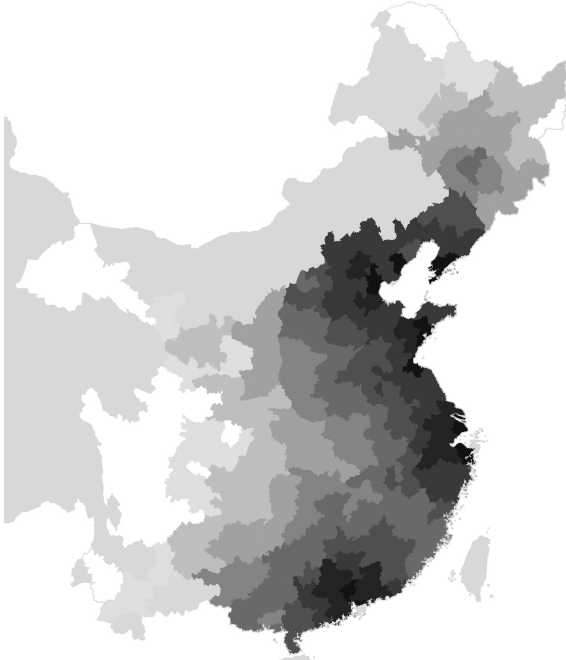


(e) Regional primates

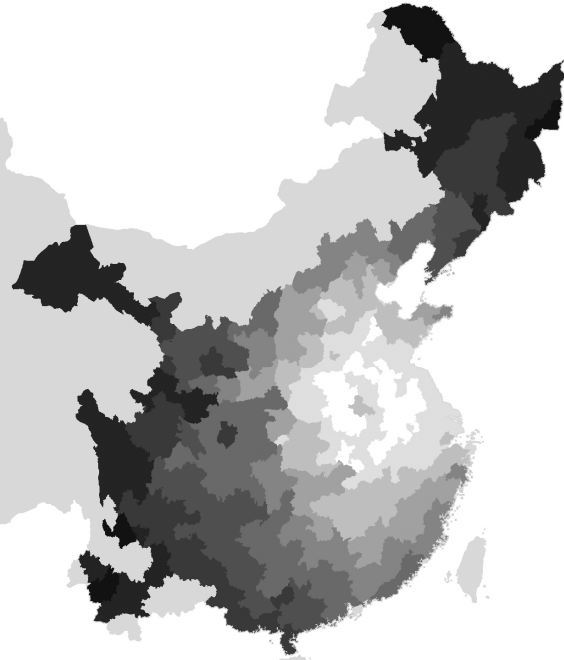
Figure 2: Geographic distributions of road measures. In all panels lighter colors indicate larger values.



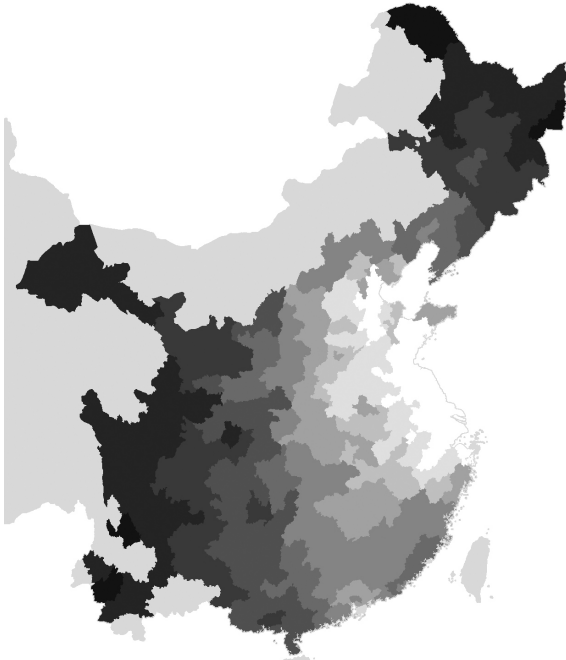
(a) 2010 Efficiency roads



(b) Travel time to port

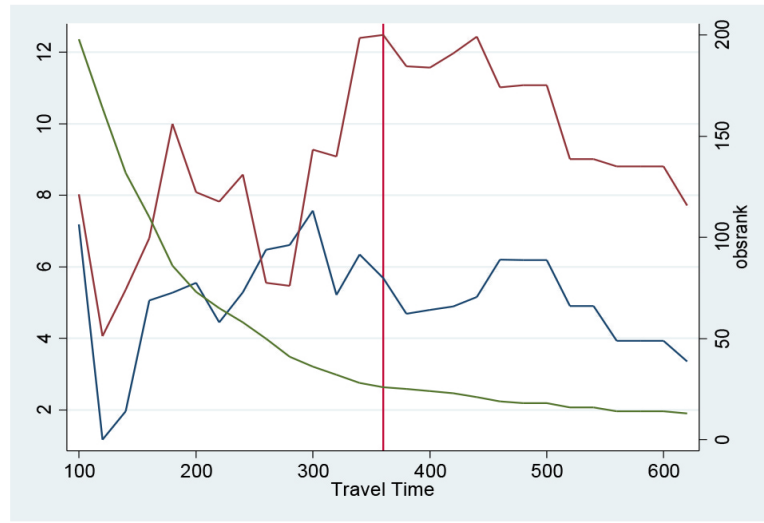


(c) Market Potential

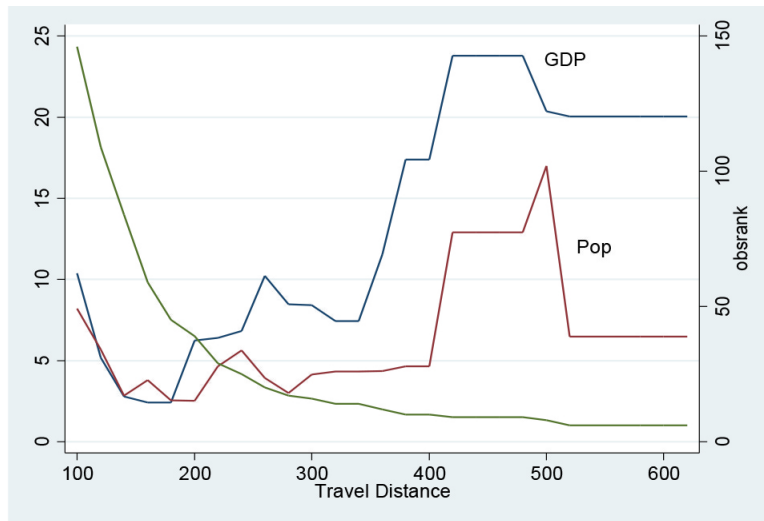


(d) Market Access

Figure 3: Determining the Number of Regional Primate Cities



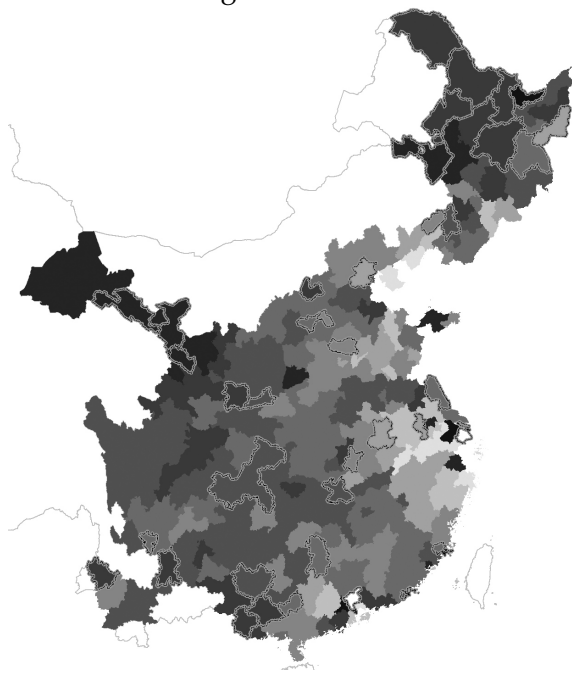
(a) Travel Time Based



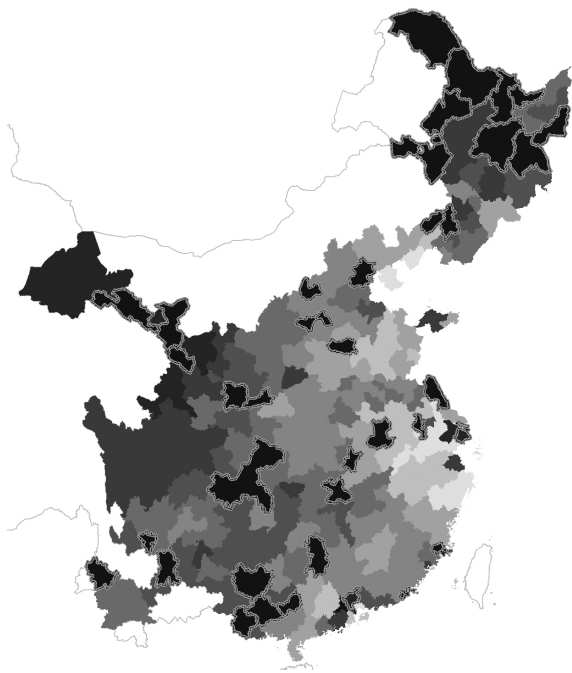
(b) Distance Based

Notes: Graph of  $\chi^2$  test statistics comparing empirical models with primate city interactions to those without for different definitions of primacy. Primacy is defined as (a) the highest population prefecture within the indicated number of minutes' drive over the 1962 road network at 90 kph or (b) km of straight line travel. The blue line uses log 2010 GDP as the outcome and the red line uses log 2010 population as the outcome. The green line indicates the count of primate prefectures for each distance threshold and definition.

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

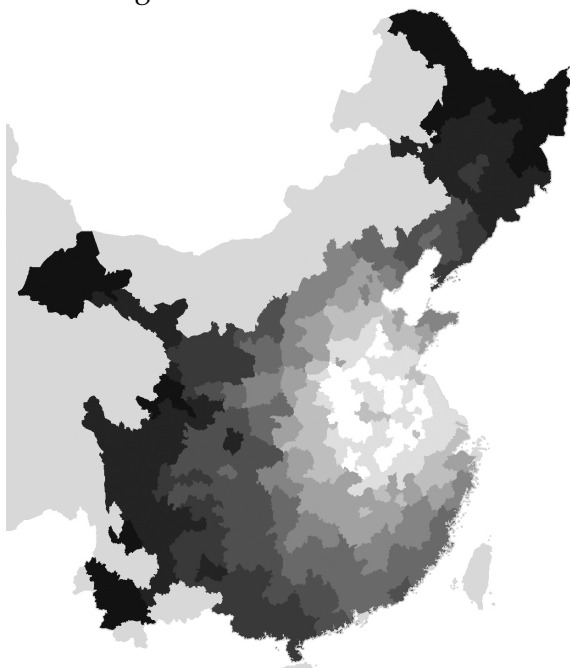


(a) Counterfactual changes in population  
no regional primate effects

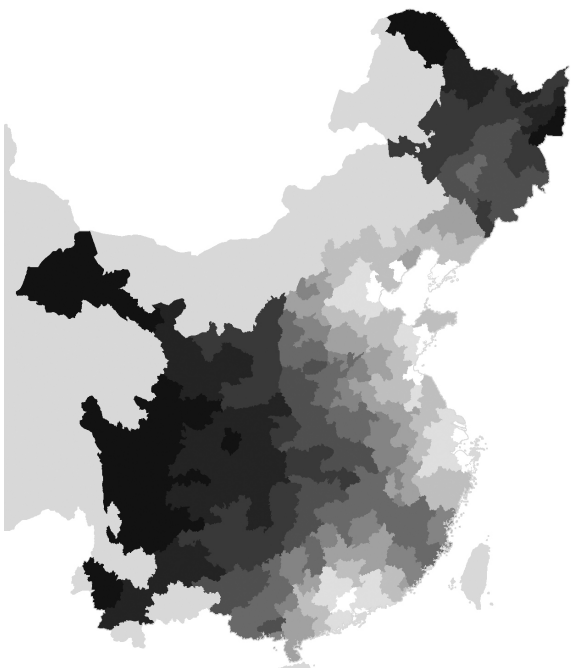


(b) Counterfactual changes in population  
with regional primate effects

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