Transport Infrastructure, Urban Growth and Market Access in China

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

September 15, 2015

Abstract

We investigate the causal effects of the massive investments in a highway network in China on economic outcomes in prefectures. We separately measure the influence of changes in access to domestic markets versus international ones. We employ two main approaches, a structural model focused on Ricardian trade forces and conventional econometric estimation focused on causal treatment effects, for which the Chinese context is a good one. While there is the usual trade-off between an approach that captures general equilibrium effects versus one focused on nailing causal effects, we find that in China the usual Ricardian forces do not dominate results. Regressions suggest improved international connections in China with its export driven growth policies are critical, and that improvements in domestic market access favor regional primate cities over others, probably due to restrictions on factor movements. These aspects are not well captured by existing structural approaches to the analysis of national infrastructure changes.

1 Introduction

Between 1990 and 2010, China constructed an extensive modern highway network including a national expressway system. We investigate the effects of this network on the distributions of economic output, population and GDP per capita across prefectures. Our investigation faces two challenges. First, highways were not randomly assigned to locations within China. Highways may have been allocated according to locations' productivity, trade potential, and/or attractiveness as a place to live. Second, output in each region depends fundamentally on output in every other region through trade linkages. As a result, highway construction may generate general equilibrium effects on trade and migration. Trading links cause the effects of a highway constructed near one particular city to percolate throughout the country. This limits our ability to assign regions to treatment and control groups, which is at the foundation of most well identified econometric analyses. The lack of a clear control group makes it difficult to distinguish between highways' effects on aggregate growth versus distribution across regions. Because there is no counterfactual for all of China, identifying the effects of the new highway system on national outcomes requires invoking and utilizing strong structural assumptions.

To investigate these issues, we implement three distinct research designs in parallel for recovering causal effects of highways on prefecture output and population. As in Donaldson & Hornbeck (2015) and Alder (2015), the first utilizes and calibrates a general equilibrium model of trade in the spirit of Eaton-Kortum (2002) [henceforth "EK"]. We modify current versions of the model to separately incorporate international and domestic market access effects. We use the model to conduct counterfactual exercises. This approach, in contrast to conventional econometric analyses, accounts directly for general equilibrium effects. However any model has a specific structure which can fail to incorporate mechanisms of first order quantitative importance. In the Chinese context, the model will fail in important ways. In addition, structural models to date do not have a component which characterizes or facilitates estimation of parameters governing the non-random assignment of roads to regions.

Our second research design involves a conventional econometric exercise in which we regress prefecture level output or population on a measure of roads within a given radius of the main city and travel time to the most accessible port. This design has two advantages. First, the road measures have a direct interpretation for policy makers. Second, we can address non-random allocation of roads by utilizing historical road networks as a source of quasi-random variation. As we discuss below, the Chinese context is particularly well suited to defend the validity of instruments. While it may be that a lot of the effects of local road expansion occur through greater local market integration (Hillberry & Hummels, 2003), it is difficult to use this regression based research design to account for system and general equilibrium effects. The regression framework only successfully uncovers slope coefficients, which may be interpreted as the relative gains or losses to one city of a marginal change in its local highway allocation. However, the framework can inform us about the nature of causal relationships between infrastructure and outcomes of interest in an equilibrium. This informs us about the validity of the assumptions in a structural model. As in Head & Mayer (2004), Redding & Venables (2004) and Hanson (2005), we alternatively estimate effects of nearby output on prefecture population and output. We focus on aggregate output reachable within a 6 hour drive, which we call "market potential". We find qualitatively similar effects of market potential as of raw infrastructure.

Our third design is hybrid of the first two. It involves using theoretically generated measures of market access from the adapted EK model as regressors in a conventional econometric analysis. While this procedure has been used in the literature (Donaldson & Hornbeck, 2015; Alder, 2015) and seems intuitive, we argue that it does not have the virtues of its two parents and suffers from their vices. By construction of the model, it is logically impossible to vary one city's market access while holding other cities' market access constant. Related, the exercise provides little information about causal effects of interventions, such as local highway construction, that are directly interpretable for policy makers. Moreover, the same identification problems of non-random assignment still arise; and, as a regression equation, its interpretation is subject to the same criticism as in the second research design. Finally, if the structural model is fully specified, a market access regression is redundant. At best it might inform us about unknown parameters of the structural model. In the Chinese context, we find that the structural model offers nothing since the regression relationships it implies do not hold, meaning that the basics of the model must be misspecified.

Reduced form estimates from the second research design indicate that expansions of regional highway networks have *negative* average effects on local population and no significant effects on local GDP. In particular, a 10 percent expansion in road length within 450 km of a prefecture city leads to an estimated 1.2 percent loss in prefecture population. In examining heterogeneity of effects, we find that regional highways are estimated to promote concentration of both output and population into regional primate cities, at the expense of other cities. This may reflect unmodeled forces from migration and capital market policies discussed below.

Unlike domestic integration, regression results indicate improved access to international ports promotes growth in GDP, population and GDP per capita for all cities. A 10 percent decline in travel time to an international port caused about 1.6 percent, 1 percent and 0.5 percent increases in GDP, population and GDP per capita respectively, with no significant differential effects for regional primate cities. The indicated welfare consequences could be very large. Facilitating better access to international markets has had a high return for cities in China in the context of export driven investment and growth policies.

In the third research design, it is possible to estimate causal effects of marginal changes in "market access", a recursive function of output in all prefectures weighted by inverse travel times, in a regression framework. Estimated coefficients on overall market access using equilibrium relationships implied by the model do not match predicted calibrated values in magnitude. More problematic, the domestic component of market access has negative estimated average effects on local growth (outweighed by the positive effects of international market access). As with the local road network regression, this suggests that the Ricardian trade forces in the EK model do not dominate determination of outcomes.

Using the structural model, we calculate counterfactual output, population and welfare associated with different road networks. We reduce expressway speeds in 2010 from 90 kph to 25 kph, as on other roads. Across a wide range of parameter values describing input cost shares and productivity dispersion across firms, we consistently find that welfare is about 5% lower in real terms under this counterfactual road network. This welfare loss is almost entirely driven by reductions in domestic market integration, a conclusion which is completely at odds with the regression based empirical evidence discussed above. Second, the model predicts population gains for cities concentrated in the denser coastal area, with losers being cities in the more sparsely populated interior who experience relatively greater losses in domestic market access, being more cut-off from dense coastal markets.

We compare these calibrated counterfactuals with reduced form counterfactuals calculated using regression estimates. Here, we impose changes in local roads and port access relevant to each city (to get differential relative effects) and then constrain absolute total population changes to be zero, with relative winners and losers city-by-city. While this of course ignores general equilibrium effects, it is suggestive of where population is predicted to relocate in the regression framework relative to the structural model. We find very different results. In particular, it is the regional primate cities scattered throughout the country which lose population, while non-primate cities gain. Moreover even without considering regional primate cities, in general there is less concentration of winners near the coast, given in the regression framework, on average, being in a dense regional market is not an advantage.

We view the different regression formulations as providing a credible description of the forces at work in suggesting how new roads have changed the spatial organization of economic activity in China, which could inform design of future structural models focused on China. Results are consistent with a context in which national and regional policies suppress domestic consumption and favor export driven growth. Our regression results suggest there are economic-political *urban hierarchy* forces influencing the movements of population, capital and firms in response to changes in regional transport networks. Export processing and other special economic zones play a big role determining specific locations of FDI and export oriented activity, induced by tax, local infrastructure and other policies; and in 2010 essentially all prefectures have export zones. In addition, overall movements of capital are constrained through the state owned banking system and movements of workers are constrained by hukou related migration policies. The more homogeneous positive estimated effects of better port access may be because capital moves fairly freely across export processing zones and migrants can move more freely across export zones into zone provided housing. However for improved domestic access, with the difficulties of long distance migration and policies affecting where production for domestic consumption expands, factors may move to a region's primate cities from other regional cities in response to better regional connections.

Our work relates to the literature in a number of ways. There are many general mechanisms through which improved market integration may promote growth, all of which cannot be tractably included in a single model. The EK framework, used by Alder (2015), Donaldson & Hornbeck (2015) and Sotelo (2015), emphasizes Ricardian gains from trade. Fajgelbaum and Redding (2014) emphasize the rise of the nontraded sector and rising demand for traded manufacturing goods for facilitating structural change and urban growth in a historical context. Topalova & Khandelwal (2011) provide evidence that lower trade costs has fostered innovation through competition in India. Lower cost access to intermediate inputs (Fujita, Krugman & Venables, 1999) and innovative ideas (Alvarez, Buera & Lucas, 2013; Buera & Oberfeld, 2014) are additional mechanisms through which trade may promote growth. While our estimated regression effects can have multiple structural interpretations and be driven by many general economic mechanisms, we organize the analysis in order to facilitate their interpretation in the contexts of the models of interregional trade.

Our evidence on the effects of reduced transport costs for enhanced integration into international markets echoes some recent literature that improved access to ports fosters local economic growth in developing country contexts. Donaldson (2014), Banerjee, Duflo, and Qian (2012) and Storeygard (2012) find that better linked hinterlands through colonial railroads in India, modern railroads in China and modern roads in Sub-Saharan Africa respectively have higher income levels. In terms of domestic interconnections, Donaldson and Hornbeck (2015) find positive effects for rural counties in the late 19th century United States, though Faber (2014) and Bird & Straub (2015) find the opposite for some rural counties served by roads in China and Brazil respectively. Sotelo (2015) finds generally positive effects of paving Peruvian roads, with some areas negatively affected because of increased competition.

This paper proceeds as follows. Section 2 discusses the unique Chinese historical and institutional context, which is well-suited for recovery of causal effects in estimation, and the data. Section 3 describes the model, simulation results, and results from estimating related market access equations. Section 4 lays out our empirical strategy and estimation results for roads infrastructure measures. It then compares results for the counterfactual of shutting down the expressway system from the road infrastructure regression with those from the model. Section 5 examines an extension based on market potential measures. Section 6 concludes.

2 Context and Data

2.1 A Brief History of Chinese Geography and Highways

The Chinese context is especially well-suited for our investigation for several reasons. Because China had essentially no limited access highways in 1990, Chinese cities have experienced large variation in expansions of internal transport networks and market access since 1990. Intercity roads had two lanes with free access and, in many places, were not even paved. Almost all goods moved by rail or river and less than 5 percent of freight ton-miles moved by road. Since then, China has undertaken massive intercity expressway construction. Construction started slowly, with only a few highways complete by 2000, but sped up so that a national scale network was essentially complete by 2010, the year for which most estimation is done. Now, well over 30% of freight ton-miles move by road.

This highway construction program has left some cities with high quality links to nearby hinterland markets and coastal ports and other cities with lower degrees of connectivity.

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

Figure 1 shows the national road networks in 1962, 1990, 2000 and 2010. We use the 1962 network to construct instruments for 2010 travel costs. These travel costs assume speeds of 25 kph on local highways and 90 kph on expressways, as is explained in more detail below. Moving forward in time, we see the national expressway system developing a little between 1990 and 1999, and most of the country is linked between 2000 and 2010.

The unique history of the Chinese transition toward a market economy is also important. While there were some market oriented reforms during the 1980s in the agricultural sector, Chinese cities remained fully planned economies until the early 1990s, with little trade in general and very little across provincial boundaries. Even agricultural markets remained highly localized, with little movement of goods across prefectures. Housing and employment were provided by local governments for a planned industrial mix, with any inter-prefectural trade flows largely proscribed in provincial capitals. While today we think of China as a free market economy in goods, although less so in factors, it is hard to ask in a time differenced sense how the highway construction altered trade and improved growth from 1990 to 2010. The starting point is not a market economy where Ricardian forces were at work. Today they are, albeit with potentially key constraints operating through factor markets. Thus we focus on a "long run equilibrium" or cross-sectional analysis for 2010.

That said, for population movements we will compare cross-sectional results with growth results, since as discussed next, China moved from a regime of very limited population movements to one where population could better respond to the new market regime and the differential opportunities offered by highway construction.

Because prefecture and city populations are outcome variables, it is important to understand the history of interregional population mobility in China. Before 2000, with the exception of a few coastal mega-cities, cities hosted few 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 un-licensed migration is no longer illegal. However, the hukou system still restricts migrant access to formal housing markets, schools, health care, and social security (Chan, 2008), restrictions that are harsher in mega-cities that are otherwise more attractive to migrants. Because of such migration restrictions, most migration in the 1990s occurred within prefectures, as farmers left the land and moved from rural to urban counties (Chan, 2005). In the 1990's, rising city productivity or demand for city output is likely to be partly reflected in rising real wages in some cities (Au & Henderson, 2006) rather than rising populations, as characterizes the urbanization process in many developing countries. The lifting of formal migration restrictions has helped raise China's urbanization rate from 37% in 2000 to almost 50% in 2010. For 2010, we approach the problem in an EK context by first assuming population has become perfectly mobile. However, we also estimate versions of GDP determination controlling for population, for which we have an instrument. In addition, we examine differential population allocation and growth effects for regional primate cities.

Despite migration restrictions, China experienced considerable migration since 1990. Some of this is rural to urban migration, where in 1990 China's population was about 29% urban, rising to around 50% in 2010. The change in urbanization has 4 components: rural areas themselves becoming urban as they industrialize, migration within provinces to more urbanized prefectures, some long distance migration to coastal cities, and intra-prefecture migration from rural parts to urban parts of the prefecture. We put aside within prefecture details here and use the prefecture as the unit of analysis to look at Han China comprehensively. Table 1 presents summary statistics showing 2010 levels and trends from 1990 to 2010 in population, GDP and GDP per capita in prefectures. Of course the big story is the enormous growth in real GDP per capita in China over 20 years.

2.2 Data

Chinese administrative geography dictates the spatial units that we use in our analysis. Provinces are broken into prefectures and prefectures into counties. Over the course of our study period, the boundaries of a number of prefectures changed, requiring painstaking work establishing county level correspondences over time to provide time consistent prefectures, which we define as of 2010. We examine 282 prefectures in Han China (about half the land area of China), omitting minority areas for data and contextual reasons, the 3 cities directly governed by provinces, and one island prefecture. Our study area covers over 85% of China's population. We use two primary types of data: tabular data from the census and city and provincial yearbooks for 1982. 1990, 2000 and 2010 and a series of large scale national road maps from 1924, 1962, 1980, 1990, 1999, 2005 and 2010.

Information on output is reported for many prefecture cities and county cities, and some prefectures, back to 1990. Since our focus is on output in 2010, we omit details of the collection of earlier output data. In 2010 we use output information from the University of Michigan's Online China Data Archive, which covers prefectures, prefecture cities and rural counties. We use 100% count National Population Census data from 1990, 2000 and 2010 to construct prefecture population and employment by industry. Individual-level 0.3 percent to 1 percent sample data drawn from 1982, 1990, 2000 and 2010 censuses enables us to construct estimates of key demographic variables at the county and urban district levels. We observe age, gender, educational attainment, occupation and sector, as well as residency (or hukou). The latter is critical to identifying migrants.

To describe the Chinese road and railroad network, we digitize a series of large scale national paper maps. We select maps from the same publisher drawn using the same projection and with similar legends to have some consistency across time. However, details of what roads are recorded and their characteristics do change over time. Using the digital maps, we calculate travel times between each pair of prefecture cities over the highway network in each year. To understand the potential importance of links to the international economy, we also calculate travel times over the road network from each prefecture city to the nearest major international ocean port, of which there were 12 in 2010. We assume travel at 25 kph on regular roads and 90 kph on highways. Our primary domestic infrastructure predictor of interest in the treatment effect oriented analysis is the log length of roads within 450 km of the main city in each prefecture. We use an efficiency units type measure that counts highways as 90/25ths of a regular intercity road to reflect their higher travel speeds. Use of this measure allows us to consider counterfactual environments in which highways are downgraded to 25 kph.

Table 2 reports statistics on key variables used in the paper, which we will refer to from time to time.

2.3 1962 Roads and Modern Highways

The econometric part of our investigation attempts to recover causal effects of 2010 highways and various measures of access to markets facilitated by these highways on contemporaneous prefecture outcomes. While we provide a detailed explanation of our main estimating equations in Section 4, any credible empirical results depend on isolation of exogenous variation in these 2010 highways. There exist a host of potential concerns in this regard. Prefectures with greater GDP and population are likely to have more resources to build highways, reflecting a reverse causal link from the outcome to highways. Moreover, higher levels of government may have provided better highway links to export nodes for prefectures specialized in export-oriented activities. In short, highway construction is likely to respond to travel and shipping demand. Picking out exogenous variation in 2010 highways requires finding a portion of such highways that were built for other reasons. As noted above, we use the 1962 road network as an instrument for the 2010 highway network and predictors of interest calculated using this 2010 network, based on the idea that 1962 roads were built for other reasons but were upgradeable to modern highways at lower cost than would be required to establish new rights of way. Areas with more vintage roads, however low quality, had lower costs of building out their highway systems. As a result, locations with more 1962 roads also had more highways in 2010.

This class of instruments is only valid if it is both a strong predictor of 2010 highways and is not correlated with variables for which we cannot control 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.[1] Because 1962 roads primarily served as connections from agricultural areas to nearby cities, we also control for urbanization with 1982 prefecture city, or urban population. We control for roughness and distance to the coast to proxy for agricultural productivity. Central city

[1] 1982 is the first year for which we have census information.

roughness enters as a separate control in order to account for productivity differences outside of agriculture. Finally, much large scale manufacturing activity historically occurred in provincial capitals. Since each province carried out most of its own economic planning, a lot of within province trade and all between province trade was directed through provincial capitals. As such, provincial capitals have different institutional and industrial histories from other cities, and we control separately for them.

Table 3 Column 1 shows the result of regressing the log of 2010 efficiency units of roads within 450 km of prefecture cities on other instruments and control variables and then the road variable counterpart in 1962 except excluding own prefecture roads in 1962.[2] In addition to being a "first stage" regression, one can think of this regression equation as representing a highway supply function. We exclude highways in the origin prefecture from the instrument 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. Results show a strong relationship between 1962 roads and 2010 highways 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 in the area, perhaps because manufactures traditionally traveled primarily by rail. Prefectures the West had less highway length nearby, as is expected given the smaller amount of economic development in these areas.

Table 3 Column 2 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 dependent variable's counterpart calculated using 1962 roads but at highway speeds. This variable has the predicted strong positive relationship, with an estimated elasticity of 0.72. 10 percent more 1962 roads within 450 km outside of the origin prefecture additionally reduce port travel time by an estimated 3 percent. Prefectures further from the coast also had longer travel times, conditional on the road network and prefecture characteristics, as may be expected.

The broad conclusion from Table 3 Columns 1 and 2 is that our instruments are strong

[2] The third instrument, which we use to pick out exogenous variation in prefecture population, is further discussed in the following subsection.

predictors of endogenous variables of interest conditional on appropriate controls and that we can separate out exogenous variation in the stock of 2010 highways nearby from exogenous variation in the travel time to the nearest international port.

2.4 Migration and Prefecture Population

Some of our analysis incorporates controls for 2010 prefecture population, in order to recover per-capita GDP effects, as we explain further in Section 4. Implementation requires isolating exogenous variation in this prefecture population. To handle the potential endogeneity of prefecture population growth, we use a migration shock instrument, following Bartik (1991) and Card (2001). The idea is to use historical migration pathways as a predictor of more recent migration. We construct this instrument by interacting the fraction of out-migrants from each province going to each prefecture between 1985 and 1990 with the total number of out-migrants from each province between 1995 and 2000. While this is not the ideal measure, as it can only mechanically predict 1995-2000 prefecture population growth, it is the best we can do with our available data. Fortunately, it is a significant predictor of 1990-2010 prefecture population growth and 2010 prefecture population, conditional on appropriate controls. The identification assumption for validity of this instrument is that 1985-1990 internal migration flows are uncorrelated with unobservables (like productivity shocks) driving 2010 prefecture GDP, conditional on control variables. Especially because the instrument is based on data from the pre-market reform period, this assumption seems plausible.

Table 3 Column 3 presents the result of this first stage regression, which can also be thought as a prefecture population supply equation. Most importantly, the coefficient on the instrument is positive as expected and statistically significant. Prefectures with greater 1982 population, provincial capitals and prefectures closer to the coast also had higher populations in 2010.

3 Model and Counterfactuals

In this section, we first develop a standard model of Ricardian gains from integration based on Eaton & Kortum (2002) that can be calibrated with our data for China in 2010. The model allows us to evaluate consequences of counterfactual road networks. In addition, it delivers useful summary measures of access to markets which we use in estimation below.

3.1 Setup and Calibration

Because our framework is very similar to Donaldson & Hornbeck's (2015) adaptation of Eaton & Kortum (2002), we only describe it in brief, with full details in the Appendix. Our primary innovation is to incorporate external trade in addition to the internal trade that is the focus of the model, as we suspect that the opening up of China to world markets disproportionately benefited cities with lower cost access to coastal ports.

Consumers have preferences $U = AX$ over over a local amenity A and the CES aggregate X over product varieties. The exogenous local amenity differs across residential locations indexed by i. Each product variety receives a Fréchet distributed productivity draw z_i at each location of production i, in which the shift parameter T_i is location specific whereas the dispersion parameter θ is common across locations. Production is Cobb-Douglas over land L, labor N and capital K such that output in each location is $Y_i = z_i L_i^{\alpha} N_i^{\gamma} K_i^{1-\alpha-\gamma}$. We use values of $\alpha = 0.1$ and $\gamma = 0.7$, based on our reading of the historical and Chinese production function literature. The magnitude of land's share in overall production, α, might arguably range from 0.05 to 0.15, but calibration results will not be sensitive to exact choices. Perfect competition ensures that income in each location is the aggregate value of trade flows to all locations, net of shipping costs. We denote domestic origin locations with i subscripts, domestic destination locations with j subscripts, and the rest of the world with x subscripts. Capital is elastically supplied to each location. Shipping costs are iceberg, in which the cost of shipping one unit of any variety between i and j is $\tau_{ij} \geqslant 1$ units of that variety.

The following system of equations describes the equilibrium.

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

$$\ln Y_i = \frac{1}{1+\theta\alpha} \ln(\kappa_1 T_i) + \frac{\alpha\theta}{1+\theta\alpha} \ln(L_i/\alpha) + \frac{\gamma\theta}{1+\theta\alpha} \ln A_i - \frac{\gamma\theta}{1+\theta\alpha} \ln U + \frac{1+\gamma}{1+\theta\alpha} \ln MA_i \quad (2)$$

$$U = A_i \frac{\gamma Y_i}{N_i} MA_i^{1/\theta} \quad (3)$$

$$\overline{N} = \sum_j N_j \quad (4)$$

(1) describes the "market access" of each location and captures two intuitive features. It

is increasing in demand, as summarized by the shipping cost weighted aggregate of GDP, but decreasing in competition, as summarized by the weighted sum of market access of all locations. Market access as detailed in the Appendix is inversely related to the price index facing consumers in a city, based on all locations' factor costs and access to those locations. We can instead express the second term in MA_i as being a function real income outside of China, writing it as $\tau_{ix}^{-\theta} \frac{Y_x}{MA_x}$, which we take to be exogenous, although exports by each city depend on τ_{ix}. We refer the first term in (1) "domestic market access" and the second term in (1) "external market access". (2) describes equilibrium GDP, which is intuitively increasing in productivity, land, the local amenity and market access. Below we consider the consequences of implementing (2) as a regression equation. The remaining equations describe utility and the population constraint. For some purposes, it is also useful to replace (2) with the equilibrium relationship between population and market access. The resulting equation is

$$\ln N_i = \frac{1}{1+\alpha\theta}\ln(\kappa_1 T_i) - \ln\gamma + \frac{\alpha\theta}{1+\alpha\theta}\ln(L_i/\alpha) - (\frac{\gamma\theta}{1+\alpha\theta}+1)(\ln A_i - \ln U) + (\frac{1+\gamma}{1+\alpha\theta}+\frac{1}{\theta})\ln MA_i. \quad (5)$$

Locations with greater market access benefit from having greater demand for their products. They also benefit from having lower prices, which draws in additional population beyond the direct effect on GDP.

We recognize that free mobility across prefectures with one national utility level U is probably a strong assumption for China. As an alternative, we consider the case in which prefecture population N_i is exogenous. In this environment, (1) and (3) continue to hold, but equilibrium output is instead given by

$$\begin{aligned} \ln Y_i \quad = \quad & \frac{1}{1+\gamma\theta+\alpha\theta}\ln(\kappa_1 T_i) - \frac{\alpha\theta}{1+\gamma\theta+\alpha\theta}\ln(\alpha/L_i) - \frac{\gamma\theta}{1+\gamma\theta+\alpha\theta}\ln\gamma \\ & + \frac{\gamma\theta}{1+\theta\gamma+\theta\alpha}\ln N_i + \frac{1}{1+\gamma\theta+\alpha\theta}\ln MA_i. \end{aligned} \quad (6)$$

When we evaluate consequences of new roads, we evaluate effects with and without population mobility.[3]

Recovery of MA_i in (1) requires information about shipping costs τ. To calculate τ_{ij},

[3]Desmet & Rossi-Hansberg (2015) models a similar environment with imperfect mobility by using auxiliary data on happiness to calibrate utility differentials across locations that can be supported in equilibrium.

we use

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

where ρ is varied between 0.5 and 2. This expression captures both the pecuniary and time (opportunity) cost of shipping. Hummels & Schaur (2013) estimate that each day in transit is equivalent to an ad-valorem tariff of 0.6-2.1 percent. Limao & Venables (2001) find that the cost of shipping one ton of freight overland for 1000 miles is about 2% of value, or about 1% per day. This expression generates the resulting target of a loss of 1.6-3.1% in value per day while also incorporating some concavity. To calculate τ_{ix}, we use

$$\tau_{ix} = 1.15\tau_{ip} \ .$$

Anderson & 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 & Venables, 2001). We treat the cost shipping from i to the nearest international port p the same as shipping to any other domestic location. Following EK, we assume $\theta = 5$, noting that calibration results are not sensitive to θ. We get the inital equilibrium value of Chinese exports E in 2010 from the national accounts.

Table 2 presents summary statistics about total MA and its components while Figure 2 depicts the spatial distribution of these on a map. All maps show prefectures ranked from highest to lowest by intensity of color, so the prefecture with highest market access has the most intense color and the lowest the least intense color. With so many ranks, it is difficult visually to distinguish those with similar ranks, but the overall pattern is clear. The maps in Figure 2 show that domestic MA is spread more smoothly over the country, as should be expected given its recursive nature. External MA is noticeably concentrated along the coast, as also should be expected. Neither has much variation across prefectures, with standard deviations for the logs of 0.04 and 0.06 respectively. Note that domestic market access is about 70% of the total in 2010. This will be important for interpreting regression coefficients on each component separately.

Taking $Y_i \ , \ E \ , \ N_i \ , \ \tau_{ij} \textit{ and } \tau_{jx}$ from the data and parameters from the literature, we recover MA_i from (1), relative A_is from (3) after normalizing initial $U = 1$ and the cluster $\varepsilon_i = \frac{1}{1+\theta\alpha}\ln(\kappa_1 T_i) + \frac{\alpha\theta}{1+\theta\alpha}\ln(L_i/\alpha)$ from (2). Once MA_i has been calculated, the real value of output in the rest of the world is calculated as $\frac{Y_x}{MA_x} = \frac{E}{\sum_j \tau_{jx}^{-\theta}\frac{Y_j}{MA_j}}$. This is all we need to evaluate effects of imposing different τ matrices in calculating counterfactual equilibria. In particular, we solve for counterfactuals using the same set of equations with $\frac{Y_x}{MA_x}$, A_i, ε_i and $\overline{N}$ as inputs calculated from the data, allowing us to solve for Y_i , E , N_i and U for new τ_{ij} and τ_{jx} calculated using counterfactual road networks.

There are interesting features to the model apart from market access allocation and division between domestic and international. In Figures 3 a and b we graph observed 2010 GDP and population, again by rank, as a reference point. While there is coastal concentration of both there are strong economic centers in the interior as well. In Figures 3 c and d, we graph the recovered values of the As and εs, again using rank-color assignments. These amenity and productivity variables essentially amount to prefecture "fixed-effects". They adjust so that the model perfectly explains the data. For the εs, higher productivity cities in China are on the coast and in traditional and newer industrial centers. For the As, it looks like high amenity places are disproportionately in the fringe areas of Han China. This in itself suggests an issue with the free mobility assumption. In (5), are these fringe areas high amenity places or places from which it is difficult to migrate, so people are trapped there at low incomes per capita and utility levels?

We can also examine the extent to which observed variation is explained by the systematic parts of the model versus these fixed effects. From (2), we see that the units of both ϵ and $\ln A$ are log income. The data imply the following regression relationships:

$$\ln Y_i = 3.210 + .390\varepsilon_i + \eta_i$$
$$(0.11)$$

and

$$\ln Y_i = 6.98 + 2.12 \ln A_i + 0.944\varepsilon_i + \widetilde{\eta}_i.$$
$$(0.39) \qquad (0.08)$$

Unsurprisingly ϵ is a powerful positive predictor of output in both equations. The R^2 of $\ln Y$ on ϵ alone is 0.77 and here the addition of logA raises it to 0.95. Inverting this, this tells us that the systematic part of the model is predicting almost precisely 5% of the variation in the data. Or, more heuristically, the Ricardian model alone is able to explain about 5% of the total variation in output. For comparison sake, city level regressions with more extensive lists of regressors often achieve R^2s of 0.6-0.8.

The corresponding regression for population is defined theoretically in (5), and is a linear function of $\ln A$ and ϵ. The analogous regressions to those above are

$$\ln N_i \quad = \quad \underset{(0.32)}{3.03 - 0.125} \ln A_i + \nu_i$$

and

$$\ln N_i \quad = \quad 7.86 + \underset{(0.46)}{0.935}\varepsilon_i + \underset{(0.09)}{3.08} \ln A_i + \widetilde{\nu}_i.$$

Here with just $\ln A$ as a covariate, the R^2 is less than 0.01. Amenities alone do a poor job of explaining population allocations and amenity values are negatively related to population. This is not surprising. Given China's internal migration restrictions, one might reasonably be suspicious of the free-mobility assumption in the model as noted in Figure 3. Adding in the ϵs which drive GDP allocations enhances the explanatory power of these fixed effects, raising the R^2 to 0.87.

3.2 Counterfactual Results

We investigate the effects of changing shipping costs in two ways. First, we examine inframarginal effects of imposing 25 kph speeds on all 2010 highways. Second, we examine the effects of increasing travel time by 5% between all locations. In some cases we distinguish between changing domestic versus external market access, as if they used different road networks.

Table 4 Panel A reports utility, GDP and exports for both classes of counterfactuals considered, under free mobility.Each quantity is expressed relative to a baseline of 1. Results in the first row show that setting all highway speeds to 25 kph is predicted to reduce utility (real income) by about 5 percent. GDP actually increases by 1.2 percent to counteract the reduction of 1.5 percent in sourcing from abroad, but prices go up more since a greater fraction of goods are now produced domestically, and at higher cost.

The second row of Panel A show the effects of increasing all pairwise travel times by 5 percent. The third row shows analogous results for increasing all domestic pairwise travel times by 5 percent. Rows two and three have almost identical results, with utility falling by 4 percent, GDP rising by about 8.35 percent and exports falling by 1.5 percent. Once again, even though GDP rises, prices rise even more to reduce welfare. The final row in

Panel A shows almost no effect of changing external trade costs on outcomes.

One message from these counterfactuals is that welfare changes are driven not only by changes in GDP but also changes in prices. This is important, especially given single equation regressions ignore such changes. The other is that the model suggests that changing access to the coast has very modest effects. This will contrast with what regression results suggest.

Table 4 Panel B reports counterfactual levels of utility given imposing 25 kph on all highways and various alternative parameter combinations. Removing the 2010 highway network consistently causes about a 5% reduction in real GDP for a wide range of reasonable values of γ, α and θ. Since the systematic part of the model accounts for only 5% of total variation in output, we would require GDP to be extremely sensitive to these parameters in order to see a big effect. The exception is changes to the scaling of trade costs ρ. To a rough approximation, doubling the scale factor doubles the cost the impact of our counterfactual scenario, and also the welfare impact.

Figure 4a shows the percent changes GDP and 4c the level changes, for the first row counterfactual reported in panel A of Table 4, again by rank-color intensity. Figure 4b shows the winners versus the losers in terms of GDP in the counterfactual. Downgrading the expressway system results in a gain for dense coastal areas and losses in the interior which now have poorer access to rich coastal markets. In Figure 4 the borders of regional primate cities are outlined in black; we will discuss regional rpimates later. Also later in Section 4.3, we compare these model counterfactuals with regression equation ones, focused on changes in population (to some degree mirroring changes in GDP).

3.3 Estimation

We have information about GDP and population for each prefecture in 2010. It is thus natural to investigate how market access influences each of these objects given the structural equations (2) and (5) using 2010 data. Each of these equations can be implemented as a regression with no assumptions needed about parameter values. Of course, with knowledge of α, γ and θ, there is no real role for estimation here. The residual ε_i is just identified, and with information about land L_i and κ_1, the local productivity draw for each location T_i can be recovered. The model is specified in such a way that given knowledge of standard parameter values, no additional information is available from estimation. Moreover, this model's parameters are probably better recovered in other ways. Nevertheless, as in Donaldson & Hornbeck (2015), we investigate the implications of estimating the relationship between market access and GDP. In addition, outside the constraints of the rest of the model, it may be valuable to investigate the causal effects of MA_i as a useful measure that summarizes connections to other markets.[4]

Estimating (2) amounts to regressing prefecture GDP on a constant, prefecture land area and market access. Note, however, that there are reasons that are both internal to and external from this model which make $\ln MA_i$ in such a regression endogenous. Internal to the model, the structural error term $\frac{1}{1+\theta\alpha}\ln(T_i)$ also appears repeatedly in MA_i. Y_i, a direct function of T_i, appears in MA_i, as does each $Y_{j\neq i}$, which themselves are functions of Y_i and so depend on T_i indirectly. That is, the key variable of interest in this regression is structurally correlated with the error term, and so OLS results in inconsistent coefficient estimates. This econometric difficulty is akin to the difficulty one faces in estimating spatial lag models. While joint estimation of (1) and (13) would resolve this problem, it is not the only reason MA is likely to be endogenous. External to the model, we also believe that the road network, and the resulting τ_{ij}, should not be taken as exogenous. Prefectures with high growth potential and strong trading links may have been more likely to receive highways than others. That is, it may be that τ_{ij} depends on T_i and T_j and τ_{ix} depends on exports from i, which is a function of T_i. In general, prefecture, provincial and national government's choices of and resources for highway construction mean the highway network is likely to be endogenous to local economic conditions and population.

Commensurate with the discussion in Section 2, our solution to this endogeneity problem is to find instruments that shift MA but are not related to local productivity. The use of IV has the additional advantage of eliminating the potential bias effects of measurement error in the prefecture land area control, which would ideally be an efficiency units measure.[5] To construct instruments, we focus in particular on the component of MA that involves connections to other prefectures τ_{ij} and to export markets τ_{ix}. Conditional on appropriate controls, as discussed above, we believe that the 1962 road network is a good instrument for the 2010 highway network. Using this idea, we instrument for log domestic MA using the km of 1962 roads within 450 km of the prefecture's main city but outside

[4] If the goal were to estimate this model's parameters, a natural course of action would be to recover them using a minimum distance estimator (or GMM incorporating a model extension to introduce stochasticity) implemented on (1), (2), (3) and (4) simultaneously. Instead, our goal is simply to determine whether this model can generate reasonable empirical predictions for China.

[5] Our controls for roughness come closer to making our control for efficiency units, but there are still likely to be unmeasured components of L_i that may be correlated with components of MA_i.

of the prefecture. We instrument for log external MA using the log of travel time to the nearest port over the 1962 network assuming a speed of 90 kph. That is, we imagine a world in which all 1962 roads were upgraded to highways. We instrument for total MA with both variables. Results in Table 3 Columns 4-6 show that first stage coefficients are significant and that each market access measure is predicted by the appropriate instrument. In addition, the 1962 road stock within 450 km of prefecture cities predicts part of external market access.

Table 5 Columns 1 and 2 report regression results in which MA is unified and broken out into domestic and external components respectively. Results in Column 1 indicate that prefectures with 10 percent greater joint domestic and international market access had about 29 percent greater GDP. This point estimate is much greater than what is predicted by the model, though we have standard errors on estimates and arguable ranges for parameters. Using our parameter values of $\gamma = 0.7$, $\theta = 5$ and $\alpha = 0.1$, the MA coefficient of $\frac{1+\gamma}{1+\theta\alpha}$ calibrates to 1.13 rather than the estimated 2.91; altering parameter values within reason keeps the calibrated number well under 2. A key result in the paper is that the external component of market access is driving the positive estimated coefficient in Column 1. The coefficient on domestic market access in Column 2 is -8.8, relative to 13.3 on external market access. Because the domestic component is about 70% of total market access, the model predicts that the coefficient on the domestic component should be about $0.7\frac{1+\gamma}{1+\theta\alpha}$ and the coefficient on the external component to be about $0.3\frac{1+\gamma}{1+\theta\alpha}$.

Given this negative domestic market access coefficient, the model is clearly not capturing something important about the data generating process for prefecture GDP. If anything, these results may indicate that faster road connections to external markets are related to regional success. However, it seems difficult to imagine that better domestic market connections may actually make regions worse off. Clearly the model is failing to capture first order other considerations. Given this, we pursue alternative empirical strategies below. However, as we investigate below, China's hukou migration restrictions may explain these results in part.

In this vein, we carry out parallel exercises to estimate (5). In the context of the free mobility variant of the model, one can recover an estimate of the Fréchet productivity dispersion parameter θ by comparing the two results. In the no mobility variant of the model, market access coefficients on population would be 0, which they are not. Results in Table 5 Columns 3 and 4 show that while overall market access is not related to prefecture population, prefectures with improved access to external markets gained population while prefectures with better domestic market access actually experienced population losses, all else equal. Again this is at odds with the Ricardian domestic gains from trade framework. We note that this second result exists only conditional on controls. Taking out controls, the coefficient on domestic market access goes to 0, indicating that locations with greater market access had better economic conditions and did relatively better in the competition for population. The role of hukou restrictions is also hinted at in the results. Explicit Chinese government policies promoting exports may help explain the positive estimated coefficient on external market access. Special economic zones, which were established to host export oriented foreign investment, have relaxed hukou restrictions relative to other areas, as explained above.

To explore equilibrium in an environment with no population mobility, we estimate (6). This amounts to estimating the same regression equation as for (2) with the addition of a control for prefecture population. Table 5 Columns 5 and 6 show these results. In these regressions, coefficients on 2010 log prefecture population (instrumented as explained above) is not significantly different from 1, consistent with the model's Cobb-Douglas production technology. The estimated MA coefficient is 2.04, which is considerably larger than the $\frac{1}{1+\gamma\theta+\alpha\theta} \approx 0.2$ predicted by the model. Once again, this result is driven by the external component of market access. The domestic component is estimated to have a negative but insignificant effect on GDP, conditional on population. This result indicates that the negative GDP effect of domestic market access is entirely driven by the negative population effect of domestic market access. GDP per capita in prefectures that are well connected to domestic markets are no lower or higher than other prefectures. However, becoming better connected to external markets is likely to be welfare enhancing.

3.4 Extensions and Alternative Measures

There are many potential explanations for which estimates from the model's main structural equations may be out of line with predictions of the model. Institutional constraints in labor and capital markets and national government export promotion policies are an explanation we have noted and will continue to explore in the empirical sections to follow. But, as with any model, there are other market mechanisms from the literature which have also been ignored.

Two extensions of the model are fairly standard and worthy of consideration. First, the existence of agglomeration economies, which could be generated by any number of microfounded mechansisms (Duranton & Puga, 2004; Rosenthal & Strange, 2004) would mean that T_i is increasing in population rather than fixed over time, contributing an additional additive term to the coefficient on $\ln N_i$ in (6) and acting as a multiplier to increase the influence of market access on GDP and population in (**??**) and (5). However, city costs are also increasing in population, which pushes in the other direction by making the denominator in the condition $U = \frac{w_i}{P_i}$ increasing in city size. Bartelme (2014) considers a model of interregional interactions that features both forces. However, the limited empirical evidence we have indicates that the elasticity of productivity with respect to population and the elasticity of city costs with respect to population are comparable (Combes, Duranton & Gobillon, 2012), meaning that these considerations may offset and thus not be quantitatively important for our purposes.

Alternative tractable models that generate similar structural relationships between local economic activity and connections to nearby markets include Redding & Venables' (2004), Hanson's (2005) and Head & Mayer's (2005) adaptations of Fujita, Krugman & Venables' (1999) "New Economic Geography" model. These earlier models begin with the assumption that each region specializes in a product and has an endogenous mass of "firms" producing different varieties using a Cobb-Douglas technology plus a fixed cost. Given the evidence of urban scale externalities and that these externalities are larger within narrow industry categories, it may be natural to think of cities as specializing in related products. Firms use immobile labor, mobile capital and a composite intermediate input imported from other locations as factors of production. Monopolistic competition delivers a fixed markup over marginal cost but 0 profits in equilibrium. The analog to market access in these studies is "market potential", given by

$$\sum_j Y_j \left(\frac{I_d}{\tau_{od}}\right)^{\sigma-1}, \tag{7}$$

where I_j is location d's price index and $\sigma > 1$ is the CES parameter of the utility function. Total income is log-linear in this market potential. We also use a market potential type measures in our empirical investigation of the importance of trade integration for urban growth.

While the theoretical framework for market access has the advantage of clarifying how better market integration can lead to local growth and provides useful guidance about estimation equations, there may be other ways in which improved market integration promotes growth. Cobb-Douglas production, Fréchet distributed productivity draws and CES preferences are all useful approximations that are likely only roughly accurate. More fundamentally, additional mechanisms may exist through which trade integration causes growth. For example, Fajgelbaum and Redding (2014) emphasize the rise of the nontraded sector and rising demand for traded manufacturing goods for facilitating structural change and urban growth. Lower trade costs may foster innovation through competition. Topalova & Khandelwal (2011) provide evidence that Indian firms became more productive with the lowering of trade barriers because of increased competition from abroad. We do not explicitly incorporate intermediate inputs to production, nor do we differentiate between different sectors in production. Because of all of these potentially important mechanisms that are not addressed theoretically, we view our proposed market access formulation as highlighting some but not all of the mechanisms through which the treatment effects of reduced interregional trade costs on GDP and population may be operating. Indeed, the divergence between calibrated and estimated coefficients, and implications about welfare consequences of new highways, points to the importance of an empirical analysis that is more agnostic about the data generating process. In our empirical work we wish to allow for maximum flexibility in the underlying data generating process and focus on recovering credible treatment effects.

4 Empirical Strategy and Results for Road Infrastructure Measures

In this section, we explain the strategies we use to recover estimates of causal effects of highway connections and trade integration on prefecture GDP, population and GDP per capita. We first show how we recover the direct effects of infrastructure. Because infrastructure measures do not have a structural dependence on GDP or population in other regions, it is straightforward to recover their treatment effects. We then present results. Finally we do the key counterfactual of shutting off the expressway system using the regression estimates and compare the results with what we find for that counterfactual using the model.

4.1 Framework

We are interested in the effects of two measures of infrastructure on outcomes. The first measure, which we denote L_{it}, describes efficiency units of roads within 450 km of the prefecture city.[6] To be consistent with the structural model and facilitate counterfactual calculations, we weight expressways by 90/25 and other roads by 1. The EK model interprets L_{it} as capturing the effects of trade integration with nearby prefectures in the region, though it could capture other mechanisms as well. Second, E_{it} denotes the travel time over the road network to the nearest international port. This measure is intended to capture the effects of integration with international markets on local economic conditions. In the context of the EK model, L and E can be thought of as reduced form measures of τ_{ij} and τ_{ix} respectively.

It is plausible that each of these infrastructure measures is partly determined by some of the same unobservables that drive outcomes of interest. To resolve this inference problem, we rely on their 1962 counterparts as instruments, as is discussed in Section 2.4. Thus, a general statement of our 'Infrastructure only' estimation problem is

$$\ln y_{it} = a + \beta \ln L_{it} + \psi E_{it} + X_i \delta + u_{it} \tag{8}$$

$$L_{it} = a_1 + \beta_1 \ln L_{i62} + \psi_1 E_{i62} + X_i \delta_1 + \eta^1_{it} \tag{9}$$

$$E_{it} = a_2 + \beta_2 \ln L_{i62} + \psi_2 E_{i62} + X_i \delta_2 + \eta^2_{it}. \tag{10}$$

In (8), y denotes prefecture GDP or population and X denotes controls. We choose the set of controls to be identical to those used in Table 5 so as to make reduced form and "structural" results more easily comparable. Because the instruments are the same, identification concerns justify this same control set. The prefecture area control performs double duty. It is structural from the model and accounts for the possiblity that more rural prefectures may have had fewer roads in 1962. Other control variables are included with the same justifications as discussed in Section 2.4. In particular, we control for variables that we have reason to believe may be correlated with an instrument and drive GDP or population.

While credible recovery of coefficients of interest β and ψ in (8) is straightforward given exogenous variation in transport measures, the interpretation of these coefficients is complicated. For example, the structural model formalizes one indirect mechanism through which L can influence y and tells us how coefficients may be heterogeneous in a

[6] We explore related measures in robustness checks.

sophisticated way. Formally, we have the following recursive system which describes β.

$$\beta = \frac{\partial y_i}{\partial MA_i}\sum_j [-\theta\tau_{ij}^{-\theta-1}\frac{Y_j}{MA_j}\frac{d\tau_{ij}}{dL_i} + \frac{\tau_{ij}^{-\theta}}{MA_j}\frac{Y_j}{MA_j}[\frac{d\ln Y_j}{d\ln MA_j} - 1]\frac{dMA_j}{dL_i}]$$

$$\frac{dMA_j}{dL_i} = \sum_k [-\theta\tau_{jk}^{-\theta-1}\frac{Y_k}{MA_k}\frac{d\tau_{jk}}{dL_i} + \frac{\tau_{jk}^{-\theta}}{MA_k}\frac{Y_k}{MA_k}[\frac{d\ln Y_k}{d\ln MA_k} - 1]\frac{dMA_k}{dL_i}]$$

As seen above, the treatment effect of nearby roads is increasing in local output's share of market access and a function of how GDP or population in each prefecture throughout the country depend on these roads. This complicated interpretation can be seen as a statement of limited external validity of these estimates. It is therefore a challenge to use such estimated effects to inform policy prescriptions. However, a more straightforward model interpretation arguably exists for ψ. In particular,

$$\frac{\partial y_i}{\partial E_i} = \frac{\partial y_i}{\partial MA_i}\frac{Y_x}{MA_x}\frac{\partial \tau_{ix}}{\partial E_i}$$

That is, given knowledge of $\frac{Y_x}{MA_x}$, ψ is informative about $\frac{\partial y_i}{\partial MA_i}\frac{\partial \tau_{ix}}{\partial E_i}$, and thus may be a useful input to welfare calculations, and (after adjustment) for application to other contexts. The exogeneity of the international location helps in simplifying the interpretation of this coefficient. Even if the market access model has its limitations, it shows how estimated effects of easier connections to domestic markets are complicated and unwieldy while estimated effects of easier connections to international markets can have a more straightforward interpretation.

Ultimately, we would like to understand the welfare consequences of the Chinese highway system. It may seem that one way to do this would be to compare coefficients for GDP and population outcomes, however care is needed here because of potentially important general equilibrium effects. For population, we can reasonably assume that treatments could not have caused the aggregate to change. China's one child policy makes it especially unlikely that highways could have promoted or dampened fertility much. However, we cannot be certain about how the highway treatments received by all prefectures in the country influenced average GDP. That is, positive estimated GDP effects may reflect positive treatment effects for GDP in more heavily treated locations and negative GDP effects in less heavily treated locations, consistent with Faber's (2014) evidence for example; alternatively there could be positive GDP effects everywhere. As with the market access regressions discussed in Section 4, we carry out a parallel analysis in which we impose constant population in regressions by controlling (and instrumenting) for 2010 prefecture population explicitly. The results of these regressions allow us to isolate variation in GDP after netting out migration effects - however we still cannot isolate the "level" effect on average GDP per capita of the highway intervention.

One message that comes out of the discussion above is the likelihood that there exist heterogeneous treatment effects of highways. In our empirical work, we focus on recovering treatment effects as a function of the importance of a prefecture in its region. We count each prefecture whose city has the largest population within a 6 hour drive over the 1962 road network at highway speeds as a regional primate prefecture. The model emphasizes how these larger sources of demand (and market access) may be expected to see bigger effects of roads, as formalized in (8).

4.2 Results

Table 6 reports coefficient estimates from (8), in which infrastructure is instrumented using 1962 counterparts. Regional infrastructure has no estimated effect on output (Column 1) and a negative estimated effect on population (Column 3). In particular, prefectures with 10 percent more road capacity nearby, measured in efficiency units, had 1.2 percent smaller populations. While these results are at odds with what would be expected in an environment with free mobility, we will examine plausible explanations based on hukou migration restrictions. Note absent controls, there are positive relationships between regional roads and both population and output. That is, higher GDP and population regions had more roads in 1962 and in 2010. However these locations gained less population than otherwise would have been expected given their underlying productivities.

Commensurate with our empirical results inspired by the structural model, we find strong evidence that better port connections led to greater local output and population. Results in Columns 1 and 3 indicate that 10% less time to an international port lead to 1.6 percent higher GDP and 1% higher population. Because this result is conditional on distance to the coast, it is driven by variation in the road network. Specification checks reveal that this result is mostly driven by variation amongst prefectures within 500 km of the coast, which is intuitive since far out prefectures are unlikely to be marginal producers for export.

As is discussed above, these reported treatment effects are likely to incorporate substantial heterogeneity across prefectues. Raising travel speeds to locations with low demand should have smaller effects from raising speeds to high demand locations. To get at this in a simple way that is informed by the Ricardian model, we investigate how treatment effects vary as a function of the importance of a prefecture in the local hierarchy. We count all prefectures as "Rank 1" if they are have the largest population within a 6 hour drive over the 1962 road network at 90 kph.

Results in Table 6 Columns 2 and 4 show that rank matters. Rank 1 fixed effects have strong negative signs, indicating that large regional cities have smaller population and GDP in 2010 than would be expected given their 1982 observables and proxies for underlying productivity. However, those rank 1 cities that got better connected to nearby areas had significantly greater GDP and population. In particular, 10% more efficiency units of roads within 450 km of rank 1 prefectures led to 4.4% higher GDP and 2.5% higher population. Remaining prefectures exhibit a negative relationship between road connections and population, with a coefficient of -0.16. That is, it seems highways caused people to migrate from other prefectures to regional primate cities. While our data does not provide much information on migration paths, we suspect that most of this migration is fairly local. Migration is less costly for moves to nearby cities since living without local hukou is feasible and engineering hukou changes from nearby prefectures is easier in some areas of the country. These results are also consistent with Faber's (2014) evidence that Chinese highways displaced economic activity from rural regions to nearby cities. We do not find any evidence that regional primate cities benefit more from faster port connections.

One potential identification concern about these results is that 1962 highways are correlated with unobservables about cities that are fixed over time. To allay this concern, Columns 5-6 of Table 6 show population results differenced between 1990 and 2010. They are almost identical to the levels results in Columns 3-4. Given the 1982 controls, one can think of the Columns 3-4 regressions as being first-differenced already. Because we have incomplete and poorly measured GDP data for 1990, we do not present 1990-2010 differenced GDP results.

Columns 7-8 of Table 6 present regression results analogous to those in Columns 1-2 with the addition of a control for 2010 prefecture population. This 2010 population control is instrumented with predicted migration flows, as is explained in Section 2.4. The reason that these results are not exactly the same as subtracting coefficients in Column 3 from those in Column 4, for example, is that here population is explicitly held constant. Therefore, all variation across prefectures predicted by roads comes through remaining components of GDP, which may include market access. Results indicate insignificantly greater per-capita GDP in prefectures with more roads built nearby, which may be driven by greater market access in these locations. However, we do find greater per-capita GDP in locations with faster connections to international ports. In particular, 10 percent faster travel to an international port increases GDP per capita by about 0.5 percent. We find no conclusive evidence that rank matters for capita GDP effects. That is, the rank effects on GDP in Column 2 appear to be driven by the effects on population in Column 4.

4.3 Reduced Form versus Model Counterfactuals

As we discussed above, there are limitations on the conclusions we can draw about counterfactuals using the regression results in isolation. Imposing a fixed national population allows us to make predictions about prefecture population changes under counterfactual road environments, albeit using estimates based on the marginal effect on any one city from the current equilibrium of a marginal infrastructure change facing that one city. However, recovering the effect on average GDP or GDP per capita in response to counterfactual road networks is even more strained, given we expect from the model that general equilibrium price effects are critical in evaluating welfare changes and there is no population adding up constraint with with to adjust single equation estimates. So in this section we focus on population changes.

Table 7 shows the results of the counterfactual exercise of shutting down the expressway system by setting expressway speeds to ports to 25kph and in counting local roads giving expressways a weight of 1 rather than 90/25. In Row 1, columns 1-3 we show results calculated using an equation that has no regional primate city distinction, while Rows 2-4 show results when there is heterogenity of effects for regional primate versus all other cities. Row 2 shows average effects across cities, while Rows 3 and 4 break out gains and losses for regional primate versus all other cities. In Columns 1 and 2, we separately examine the effects of just altering regional road counts or domestic access and just altering driving times to ports, with no adjustment for a national population constraint (because there is no clear way to adjust components).

Consistent with Table 6, in Rows 1 and 2, on average cities gain with reduced local access and lose because of reduced port access. In Rows 1 or 2, summing the average effects of the two components leaves a small net loss. In column 3 we show both operating together with all city populations adjusted by the same proportion so the net overall change is zero.

The numbers in parentheses are the standard deviations of changes and show the degree of "churning". In Rows 3 and 4, we break out the changes for regional primate versus all other cities. All types of cities suffer from reduced access to the coast, with relative variation. Regional primate cities suffer from reduced local market access, while other cities reclaim population from the regional primates under the counterfactual.

How do results differ from the model counterfactual? Model results are in Column 4 where there is no heterogeneity of regional primates in the model. Still in Rows 3 and 4 we can show what the model predicts for regional primates versus all other cities. There as expected regional primates are similar in responses, as all other cities, unlike the regression model with heterogeneity. There are two other points of comparisons. First in Row 1 are standard deviations of changes, where under the model the degree of churning, or the standard deviation, is similar to the regression model, with a modest reduction in churning. Second, are maps of the spatial patterns of changes. These are in Figure 5.

Figure 5a shows the model predictions as to the relative gain in population done in percent changes, using again the rank-color scheme. In Figure 5 the borders of regional primate cities are again outlined in black. As with GDP in Figure 4, winners are on the coast and near coast regions in the dense part of the country. The most intense gainers are on the Beijing-Shanghai axis and their hinterlands. Figure 5b shows a dichotomous split: gainers in population in blue and losers in red. Figures 5c and 5d show intensity of gains and then winners versus losers respectively, for the Table 6 Column 3 specification. Figures 5e and 5f repeat this for the Table 6 Column 4 coefficients where regional primate cities experience differential effects. Even without distinguishing regional primate differentials, in 5c and 5d we can see that there is less gain for the dense coastal areas favored in part a; and there are now interior gainers, who have lower domestic market access to begin with. Note in the regression counterfactuals (and model) all cities lose because of reduced market access, so differentials are driven by domestic considerations. In Figures 5e and 5f we see the role of regional primate cities, who are the intense losers from reduced regional networks, while other cities in each of their hinterlands gain, resulting in a spread of gainers throughout the country. The contrast between model predictions in 5b (or a) and 5f (or e) is pretty stark.

5 Extension to Market Potential Measures

One difficulty with looking at the direct effects of infrastructure is the likelihood of heterogeneous effects as a function of which locations the treatment highways are connecting. The model clarifies how reductions in transport costs have greater impacts on economic outcomes if they are between places with goods to trade. This observation leads us to consider measures of market potential like (7) as alternative predictors, with exogenous variation in this market potential achieved through exogenous road upgrades. In particular, we consider aggregate output reachable within a 6 hour drive over the road network as our primary market potential measure, denoted

$$MP_i = \sum_{j \neq i} Y_j 1\left(\text{ hours of travel time }_{ij} < 6 \right) .$$

In principle, one may like to choose a gravity measure as in (7) or a nonparameteric version thereof, and we experimented with those getting similar results. However, the challenges associated with estimating treatment effects of market potential measures are sufficiently large that we only have the power to use one at a time. Our MP_i measure has the advantage of being easy to quantify for policy evaluation purposes.

While market potential is a theoretically appealing way to measure the extent of a transportation network, an examination of the relationship between market potential and economic outcomes presents formidable econometric and conceptual challenges. The crux of the difficulty is that output is a function of output in nearby locations. Therefore, any unobserved components of output are also spatially correlated, and the independent variable of interest is thus correlated with the error term by construction. For GDP as an outcome, we have an estimation equation like

$$\ln y_i = s + \lambda \ln MP_i + \phi E_i + X_i \mu + \nu_i. \tag{11}$$

Because the only source of variation in "market potential" available from external markets is the access to export nodes, we maintain the same measure for connection to external markets, E_i as above. Using $MP_i = \sum_j e^{s+\lambda \ln MP_j + \phi E_j + X_j \mu} e^{\nu_j} 1(\text{hours of travel time}_{ij} < 6)$, we see that $\ln MP_i$ is correlated with v_i by construction. Assuming that we know that this is the true data generating process for $\ln y_i$ (such that there are no heterogeneous coefficients and that v_i is orthogonal to other right hand side variables), there are established techniques to recover parameters of this spatial lag model (Kelejian & Prucha, 2010). However, we would like to allow for flexibility in model specification such that these standard methods will not apply here. Gibbons, Overman & Pattacchini (2015) discuss the pitfalls of taking spatial lag estimation too seriously.

Our solution is to make use of a truly exogenous component of $\ln MP_i$ as an instrument - the km of 1962 roads within 450 km of each prefecture city but outside of the prefecture. Results in Table 3 show that this is a strong predictor of market potential. The set of control variables X is chosen exactly the same as above for the same reasons. Because the instruments are exactly the same, the justifications of appropriate control variables is exactly the same. This way of setting up the empirical work has the additional advantage of making results in Tables 5, 6 and 7 directly comparable, as they only differ by their dependent variables.

Note that this description of market potential tries to capture the idea that trade within six hours drive is cheap, and beyond that is prohibitively expensive. This is broadly consistent with observation in the US, where the preponderance of manufactured goods are shipped less than this distance (Hillberry & Hummels, 2005). However, one can imagine that the relationship between a prefecture outcome and connectivity to nearby places is more nuanced. In fact it is straightforward to generalize to allow for the effect of market potential to vary with driving distance, e.g., with 3, 6, or 9 hours drive. Practically, identification challenges arise when we try to do this. We do not have sufficient first stage power to achieve separate exogenous variation in market potential in different time bands. This means that market potential results should not be interpreted as strictly applying to 6 hours' driving time, but instead to the amount of economic activity reachable by road in some sense.

Table 8 reports estimated effects of increasing GDP accessible within a 6 hour drive alongside port access effects. These results are quite similar to the direct infrastructure results. In particular, we find no direct effects of market potential on GDP and negative effects on population. Prefectures with 10% greater market potential are estimated to have 7.8 percent lower population. Port access matters the same as in the raw infrastructure regressions in Table 6, as should be expected given that instruments can separate out exogenous variation in port access from market potential. As with the infrastructure results, we find that rank matters for the effects of market potential but not for port access.

At first glance, it might seem remarkable that results tell the same story for both infrastructure and market potential. Indeed, the model inspired market access regression results also give the same impression. Prefectures that became better connected to external markets experienced GDP, population and GDP per capita growth. Prefectures that became better connected to nearby areas did no better in terms of GDP and lost population, resulting in potentially small GDP per capita gains. Lots of this relates to diversion of population from rural prefectures to nearby primate cities.

The econometric explanation for the similarity in these results is that the instruments in all three cases are the same. That is, the variation in each of these three classes of variables that is being used to identify coefficients is the variation induced by the 1962 road network. Therefore, differences between coefficients in Tables 5, 6 and 8 must be fully accounted for by differences in first stage rather than reduced form relationships. The exogenous variation in nearby roads efficiency units, market potential and market access is thus highly correlated by construction, thereby generating similar results.

While we recover coefficients that are not significantly different from 0 on GDP and GDP per capita, we emphasize that this does not mean that building the highway network in China resulted in no GDP effects, only that any GDP effects impacted prefectures equally. Because there is only one China in our data set, we have no statistical power to recover such potential level effects.

6 Conclusion

In this paper we apply the workhorse Eaton-Kortum model to analyze the impact of the construction of the expressway network in China on the output and population of prefectures. We find that the Ricardian domestic trade forces that are central to the EK model have not been important in China, even from regressions based directly on that model. Rather, there are two features which arise in regression equations, which seem central to the process. First is the role of access to coastal ports which are a key driver in regressions, which might be expected in a country with export driven growth as a policy. Second are the hierarchy forces at work influencing outcomes. Domestic development spurred by highways is focused on regional primate cities, at the expense of other cities in their hinterlands. We speculate that this pattern may be driven by hukou and capital market policies channelling resources for domestic development to regional primates. On the other hand, for access to ports there is no differential in effects by place in the urban hierachy, in a context where resources flow pretty freely across the now ubiquitous export processing zones.

A Derivation of Model Equilibrium Conditions

The marginal production cost of a unit of a variety produced at location i is $\frac{q_i^a w_i^\gamma r^{1-\alpha-\gamma}}{z_i}$, where z_i is productivity, q_i is land rent, w_i is the wage. This Cobb-Douglas form delivers $\gamma Y_i = w_i N_i$ and $\alpha Y_i = q_i L_i$, in which Y is total output, N is labor and L is land.

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

$$X_{ij} = \kappa_1 T_i (q_i^a w_i^\gamma)^{-\theta} \tau_{ij}^{-\theta} \frac{Y_j}{CMA_j}. \tag{12}$$

In (12), Y_j is destination income or GDP, $\kappa_1 = [\Gamma(\frac{\theta+1-\sigma}{\theta})]^{-\theta/(1-\sigma)} r^{-(1-\alpha-\gamma)/\theta}$ where σ is the elasticity of substitution parameter in preferences, and CMA_j denotes "consumer market access", which summarizes how accessible competing markets are for provision of goods to d. Adding up the value of all flows into China from this expression, we have $I = \kappa_1 T_x (q_x^a w_x^\gamma)^{-\theta} \sum_d \left[\frac{Y_d}{CMA_d} \tau_{xd}^{-\theta}\right]$. In these expressions,

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

From (12), we see that more productive and lower cost origins ship more everywhere, more is shipped to nearer destinations with lower values of τ_{ij}, to those destinations with more income, and to those destinations with less competition from other locations. If θ is higher, that means less productivity dispersion, so it is less likely that any given origin is going to have a comparative advantage in producing as many varieties. CMA_j is closely related to the price index P_j for location d. In particular, it aggregates the marginal production costs across locations that supply goods to d. Prices are lower, and consumer market access is higher, in locations that are better linked to other productive locations.

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

$$Y_i = \kappa_1 T_i (q_i^a w_i^\gamma)^{-\theta} \left(\sum_j \tau_{ij}^{-\theta} \frac{Y_j}{CMA_j} + \tau_{ix}^{-\theta} \frac{E}{\sum_i \kappa_1 T_i (q_i^a w_i^\gamma)^{-\theta} \tau_{ix}^{-\theta}} \right) \tag{13}$$

The second term within brackets is derived by setting Chinese exports E equal to the sum of the value of all trade flows to x and can be rewritten as $\tau_{ix}^{-\theta} \frac{Y_z}{CMA_x}$. We see that GDP is decreasing in local production costs and increasing in destinations' GDP. If nearby destinations have greater consumer market access, total income is reduced because of greater nearby export competition. Denoting the term in brackets as "firm market access" FMA_i, and inverting (13) to substitute for $\kappa_1 T_i (q_i^a w_i^\gamma)^{-\theta}$ within FMA_i, and substituting for $\kappa_1 T_x (q_x^a w_x^\gamma)^{-\theta}$ in CMA_j using aggregate import flows, we have the following equations, which reveal that $FMA_i = CMA_i = MA_i$ if imports equal exports.

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

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

With free mobility, it must be the case that the real wage is equalized everywhere, or $A_i \frac{w_i}{P_i} = U => w_i = \frac{U}{A_i} MA_i^{-1/\theta}$. Therefore, we have the following equilibrium relationship between population, output and market access at each location: $N_i = \frac{\gamma Y_i}{w_i} = \frac{A_i \gamma Y_i}{U MA_i^{-1/\theta}}$. Substituting for q_i and w_i in (13), we derive equilibrium output in each location: $\ln Y_i = \frac{1}{1+\theta\alpha} \ln(\kappa_1 T_i) + \frac{\alpha\theta}{1+\theta\alpha} \ln(L_i/\alpha) + \frac{\gamma\theta}{1+\theta\alpha} [\ln A_i - \ln U] + \frac{1+\gamma}{1+\theta\alpha} \ln MA_i$

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

References

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

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

Anderson, James E. & Eric van Wincoop (2004) "Trade Costs," Journal of Economic Literature, 42:3, 691-751.

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

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

Bartelme, Dominick. (2014) "Trade Costs and Economic Geography: Evidence from the U.S." manuscript.

Bartik, T., 1991. Who Benefits from State and Local Economic Development Policies? Kalamazoo (MI):W.E. Upjohn Institute for Employment Research.

Baum-Snow, N. (2007) "Did highways cause suburbanization?" Quarterly Journal of Economics 122:2, 775–805.

Baum-Snow, N. (2015) "Urban Transport Expansions, Employment Decentralization, and the Spatial Scope of Agglomeration Economies," manuscript.

Bird, Julia & Stephane Straub (2015) "The Brasilia Experiment: Road Access and the Spatial Pattern of Long-Term Local Development in Brazil," manuscript.

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

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

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

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

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

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

Desmet, Klaus, David K. Nagy and Esteban Rossi-Hansberg (2015) "The Geography of Development: Evaluating Migration Restrictions with Coastal Flooding," manuscript.

Donaldson, D. (2014) “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure,” American Economic Review, forthcoming.

Donaldson, D. and R. Hornbeck (2015) “Railroads and American Economic Growth: A ‘Market Access’ Approach,” Quartely Journal of Economics, forthcoming.

Duranton, G., P. M. Morrow, and M. A. Turner (2014) “Roads and trade: Evidence from U.S. cities,” Review of Economic Studies, 81:2,681-724.

Duranton, G. and M. A. Turner (2011) “The fundamental law of road congestion: Evidence from U.S. cities,” American Economic Review, 101:6, 2616-2652.

Duranton, G. and M. A. Turner (2012) “Urban growth and transportation,” Review of Economic Studies, 79, 1407-1440.

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

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

Faber, B. (2014) “Trade Integration, Market Size, and Industrialization: Evidence from China’s National Trunk Highway System,” Review of Economic Studies, forthcoming.

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

Fajgelbaum, P. and A.K. Cosar (2014) "Internal Geography, International Trade, and Regional Specialization," manuscript.

Fujita, Krugman and Venables (1999) A Spatial Knowledge Economy.

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

Gibbons, Steve, Henry Overman & Eleonora Patacchini (2014) "Spatial Methods" in Handbook of Regional and Urban Economics Vol. 5, Duranton, Henderson & Strange Eds., Elsevier.

Hanson (2005) "Market Potential, Increasing Returns, and Geographic Concentration,” Journal of International Economics, 67(2005): 1-24.

Head & Mayer (2004) "Market potential and the location of Japanese investment in the European Union.” Review of Economics and Statistics.

Henderson, J. V. (1974) "The sizes and types of cities," The American Economic Review, 640-656.

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

Henderson, J. V., A. Storeygard and D. Weil (2012) “Measuring Economic Growth from Outer Space,” American Economic Review.

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

Hummels, David (2001) "Time as a Trade Barrier," manuscript.

Hummels, David & Georg Schaur (2013) "Time as a Trade Barrier," American Economic Review 103, 2935-59.

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

Lee, K.S. (1982) "A model of intra-industry employment location: an application to Bogota Columbia," Journal of Urban Economics, 12, 263-279.

Lee K.S. and S.-C. Choe (1990) "Changing location patterns of industries and urban decentralization policies in Korea," in J. K. Kwon, Editor, Korean Economic Development, Greenwood, New York.

Lewis, E. (2011) "Immigration, skill mix, and capital skill complementarity." The Quarterly Journal of Economics, 126(2), 1029-1069.

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

Lucas, R.E. (1988) "On the mechanics of economic development," Journal of Monetary Economics, 22, 3-42.

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

Redding, S. and A.J. Venables (2004) "Economic geography and international inequality," Journal of International Economics, 62, 53-82.

Roberts et al. (2012) RSUE

Rosenthal, S. and W. Strange (2004) "Evidence on the nature and sources of agglomeration economies," in Handbook of regional and urban economics, 4, 2119-2171.

Rossi-Hansberg, E., & Wright, M. L. (2007). Urban structure and growth. The Review of Economic Studies, 74(2), 597-624.

Sotelo, Sebastian. (2015) "Domestic Trade Frictions and Agriculture," manuscript.

Storeygard, A. (2013) "Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa," World Bank Policy Research Working Paper, (6444).

Topalova, P., & Khandelwal, A. (2011). Trade liberalization and firm productivity: The case of india. Review of economics and statistics, 93(3), 995-1009.

Table 1: GDP and Popoulation Spatial Distribution and Growth

Means with Standard Deviations in Parentheses

Prefecture Outcomes

	1990	2010	Change 1990-2010
log GDP	3.78	6.82	3.04
	(0.77)	(0.94)	(0.44)
log Population	14.95	15.09	0.14
	(0.65)	(0.66)	(0.22)
Per-Capita GDP (millions)	16.50	303.2	286.6
	(12.70)	(189.6)	(181.5)

Notes: Means and standard deviations are reported for the 282 prefectures in our data that are comprised of more than one county or urban district.

Table 2: Predictors and Instruments

log 2010 Road Efficiency Units within 450 km	10.72
	(0.40)
log Road Time to Nearest	5.86
Port	(1.31)
log Total Market Access	6.52
	(0.04)
log Domestic Market	6.23
Access	(0.04)
log External Market	5.13
Access	(0.06)
log GDP Within 6 hour drive	9.95
	(1.30)
log 1962 Roads within 450 km	9.39
outside of prefecture	(0.29)
log Road Time to Nearest	6.06
Port, 1962 (fast)	(1.42)
Rank 1 Prefecture Indicator	0.09
	(0.29)

Market access variables are calculated as explained in the text.

Table 3: First Stage Regressions

	log 2010 Road Effiency Units within 450 km	log 2010 Time to Nearest Port	log 2010 Prefecture Population	log 2010 Market Access	log 2010 Domestic Market Access	log 2010 External Market Access	log 2010 GDP Within 6 Hours
Instruments							
log 1962 Roads within	1.05***	-0.30**	-0.056	0.081***	0.088***	0.059***	1.50***
450 km, Excl own Pref	(0.04)	(0.13)	(0.058)	(0.0076)	(0.0083)	(0.0081)	(0.24)
log 1962 Time to Nearest	-0.016*	0.72***	-0.025	-0.0029**	-0.00073	-0.0093***	-0.054
Port Given Road Upgrades	(0.01)	(0.072)	(0.019)	(0.0014)	(0.0016)	(0.0015)	(0.04)
Migration Instrument	1.8e-07*	-8.7e-07**	1.2e-06***	2.3e-08**	2.5e-08**	1.6e-08*	4.6e-08
	(8.72e-08)	(4.1e-07)	(3.3e-07)	(9.6e-09)	(1.2e-08)	(9.6e-09)	(2.8e-07)
Controls							
log Prefecture Area, 2005	-0.079***	-0.060	-0.029	-0.011***	-0.014***	-0.0046	-0.51***
	(0.02)	(0.053)	(0.026)	(0.0034)	(0.0038)	(0.0036)	(0.11)
log Central City Area, 1990	0.0099	0.031	-0.039*	-0.00083	-0.00081	-0.00082	-0.022
	(0.01)	(0.047)	(0.022)	(0.0020)	(0.0024)	(0.0020)	(0.06)
log Central City Population,	-0.039**	0.012	0.011	-0.0055**	-0.0062**	-0.0038	-0.0061
1982	(0.01)	(0.062)	(0.023)	(0.0025)	(0.0028)	(0.0027)	(0.08)
log Central City Roughness	-0.0036	0.047	-0.0070	0.00045	0.00051	0.00025	-0.016
	(0.01)	(0.049)	(0.014)	(0.0015)	(0.0017)	(0.0017)	(0.04)
log Prefecture Roughness	-0.020**	-0.037	0.0020	-0.0021*	-0.0021	-0.0021*	-0.032
	(0.01)	(0.033)	(0.012)	(0.0013)	(0.0015)	(0.0012)	(0.03)
Provincial Capital	0.035	0.12	0.26***	0.0021	0.0044	-0.0040	-0.16
	(0.04)	(0.13)	(0.041)	(0.0053)	(0.0064)	(0.0052)	(0.14)
log Prefecture Population,	0.080***	0.019	0.82***	0.014***	0.017***	0.0053	0.58***
1982	(0.02)	(0.074)	(0.045)	(0.0038)	(0.0042)	(0.0044)	(0.13)
Share Prefecture Population	-0.83***	-0.94	-0.48	-0.044	-0.071	0.037	-0.60
with High School, 1982	(0.31)	(0.98)	(0.42)	(0.045)	(0.050)	(0.044)	(1.22)
Share Prefecture Population	-0.24	-0.45	-0.52*	0.011	0.00013	0.044**	0.73
in Manufacturing, 1982	(0.18)	(0.59)	(0.26)	(0.020)	(0.024)	(0.019)	(0.57)
log km to Coast	0.00017	0.062**	-0.026**	-0.0040***	-0.0031**	-0.0066***	-0.023
	(0.01)	(0.029)	(0.013)	(0.0012)	(0.0013)	(0.0017)	(0.04)
West Region	-0.26***	0.071	-0.020	-0.032***	-0.023***	-0.057***	-0.97***
	(0.03)	(0.087)	(0.042)	(0.0051)	(0.0055)	(0.0058)	(0.15)
East Region	-0.014	-0.16	-0.050	0.012***	0.0040	0.038***	0.37***
	(0.02)	(0.10)	(0.039)	(0.0033)	(0.0038)	(0.0039)	(0.11)
Constant	0.73**	4.06***	4.25***	5.80***	5.42***	4.68***	-7.01***
	(0.36)	(1.51)	(0.81)	(0.081)	(0.087)	(0.087)	(2.47)
R-squared	0.88	0.88	0.92	0.81	0.75	0.88	0.76

Notes: Each column is a separate representative first stage regression. Each regression includes 282 observations.

Table 4: Results from the Quantitative Model
Means Across Prefectures Relative to Baseline of 1

	Free Mobility		
	Utility	GDP	Exports
Panel A: Counterfactual Results			
Set All Highway Speeds to s to 25 kph	0.948	1.012	0.985
Increase all travel minutes by 5 percent	0.960	1.082	0.985
Increase domestic travel minutes by 5 percent	0.960	1.085	0.985
Increase travel minutes to port by 5 percent	1.000	0.998	1.000

Panel B: Robustness for Reducing all Highway Speeds to 25 kph Given Free Mobility

theta	alpha	gamma	rho	Utility	GDP	Exports
3	0.1	0.7	1	0.949	0.983	1.003
10	0.1	0.7	1	0.950	0.990	1.034
5	0.05	0.7	1	0.947	0.986	1.013
5	0.15	0.7	1	0.950	0.984	1.011
5	0.1	0.6	1	0.945	0.981	1.009
5	0.1	0.8	1	0.951	0.988	1.014
5	0.1	0.7	0.5	0.972	0.992	1.006
5	0.1	0.7	2	0.909	0.974	1.021

Notes: Each row shows the average of the object in each column header as a result of imposing the counterfactual listed at left. Each counterfactual in Panel A uses parameter values $\alpha=0.1$, $\gamma=0.7$, $\rho=1$, $\theta=5$. Shipping speeds are 25 kph on ordinary roads and 90 kph on highways. Exports in 2010 were 107022.8 million RMB.

Table 5: Market Access Regressions

	log Prefecture GDP, 2010		log Prefecture Pop, 2010		log Prefecture GDP, 2010	
log Market Access	2.91*		0.63		2.04*	
	(1.61)		(0.93)		(1.24)	
log Domestic Market Access		-8.79*		-6.84**		-1.20
		(4.59)		(3.42)		(2.10)
log External Market Access		13.3**		8.54*		3.82*
		(5.73)		(4.62)		(2.13)
lcensuspop2010_pref					1.19***	1.11***
					(0.12)	(0.13)
log Prefecture Area, 2005	0.0079	-0.093	-0.034	-0.10**	0.045	0.019
	(0.071)	(0.092)	(0.036)	(0.049)	(0.059)	(0.064)
log Central City Area, 1990	-0.083*	-0.10*	-0.023	-0.035	-0.056	-0.062
	(0.048)	(0.058)	(0.025)	(0.031)	(0.037)	(0.040)
log Central City Population,	0.12**	0.11*	0.033	0.028	0.082*	0.083
1982	(0.056)	(0.068)	(0.028)	(0.032)	(0.048)	(0.051)
log Central City Roughness	-0.059*	-0.054	-0.00068	0.0028	-0.058**	-0.057**
	(0.032)	(0.038)	(0.013)	(0.017)	(0.026)	(0.028)
log Prefecture Roughness	-0.013	-0.0060	0.0051	0.0097	-0.019	-0.017
	(0.026)	(0.032)	(0.011)	(0.014)	(0.021)	(0.022)
Provincial Capital	0.60***	0.73***	0.33***	0.41***	0.21**	0.27***
	(0.11)	(0.15)	(0.045)	(0.086)	(0.090)	(0.099)
log Prefecture Population,	0.51***	0.63***	0.81***	0.88***	-0.44***	-0.35***
1982	(0.11)	(0.11)	(0.074)	(0.051)	(0.11)	(0.12)
Share Prefecture Population	1.03	-0.41	0.14	-0.82	0.86	0.51
with High School, 1982	(0.98)	(1.06)	(0.51)	(0.55)	(0.68)	(0.74)
Share Prefecture Population	2.56***	1.47*	-0.051	-0.78	2.61***	2.33***
in Manufacturing, 1982	(0.49)	(0.78)	(0.24)	(0.55)	(0.35)	(0.39)
log km to Coast	-0.059*	0.032	-0.039**	0.022	-0.013	0.0078
	(0.035)	(0.046)	(0.020)	(0.025)	(0.025)	(0.031)
West Region	0.013	0.47*	0.030	0.33	-0.027	0.099
	(0.12)	(0.27)	(0.058)	(0.20)	(0.093)	(0.12)
East Region	0.23***	-0.28	0.021	-0.32*	0.21***	0.075
	(0.084)	(0.24)	(0.041)	(0.19)	(0.065)	(0.100)
Constant	-20.8**	-16.0	-0.86	1.66	-19.1**	-17.8**
	(10.0)	(10.9)	(5.46)	(5.67)	(7.90)	(7.56)
First stage F	68.2	20.8	68.2	20.8	8.67	10.7

Table 6: Infrastructure Regressions

	log Prefecture GDP, 2010		log Prefecture Pop, 2010		D_censuspop9010_pref		log Prefecture GDP, 2010	
log Road Eff. Units within	-0.029	-0.13	-0.12**	-0.16**	-0.13***	-0.16***	0.100	0.056
450 km of Prefecture City	(0.13)	(0.14)	(0.06)	(0.07)	(0.04)	(0.05)	(0.11)	(0.11)
X Rank 1 Prefecture		0.44**		0.25***		0.23***		0.16
		(0.19)		(0.09)		(0.07)		(0.16)
log Driving time to nearest	-0.16**	-0.18**	-0.10*	-0.11*	-0.069**	-0.075**	-0.047*	-0.051*
international port	(0.07)	(0.08)	(0.05)	(0.06)	(0.03)	(0.03)	(0.03)	(0.03)
X Rank 1 Prefecture		0.080		0.032		0.0096		0.043
		(0.08)		(0.05)		(0.03)		(0.05)
log Prefecture Population, 2010							1.09***	1.13***
							(0.14)	(0.12)
Rank 1 Prefecture		-5.16**		-2.82**		-2.43***		-1.97
		(2.26)		(1.15)		(0.86)		(1.85)
log Prefecture Area, 2005	-0.043	-0.057	-0.057*	-0.066**	-0.051*	-0.051*	0.019	0.018
	(0.06)	(0.07)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.06)
log Central City Area, 1990	-0.10**	-0.092*	-0.033	-0.031	-0.025	-0.024	-0.064*	-0.057
	(0.05)	(0.05)	(0.03)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)
log Central City Population,	0.12**	0.10*	0.033	0.028	0.031*	0.028*	0.080	0.073
1982	(0.06)	(0.06)	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)
log Central City Roughness	-0.049	-0.053	0.0045	0.0040	0.0043	0.0020	-0.054**	-0.057**
	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.03)	(0.03)
log Prefecture Roughness	-0.022	-0.028	-0.00022	-0.0038	0.0027	0.00015	-0.022	-0.024
	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Provincial Capital	0.65***	0.69***	0.36***	0.38***	0.26***	0.28***	0.26***	0.26***
	(0.11)	(0.11)	(0.05)	(0.06)	(0.04)	(0.04)	(0.10)	(0.09)
log Prefecture Population,	0.56***	0.56***	0.83***	0.82***	-0.095***	-0.11***	-0.34***	-0.37***
1982	(0.09)	(0.09)	(0.05)	(0.05)	(0.03)	(0.03)	(0.12)	(0.11)
Share Prefecture Population	0.49	0.58	-0.25	-0.26	-0.38	-0.44	0.76	0.88
with High School, 1982	(0.92)	(0.93)	(0.42)	(0.44)	(0.34)	(0.33)	(0.70)	(0.70)
Share Prefecture Population	1.96***	1.94***	-0.49	-0.51	-0.10	-0.10	2.49***	2.52***
in Manufacturing, 1982	(0.57)	(0.58)	(0.36)	(0.37)	(0.22)	(0.22)	(0.37)	(0.37)
log km to Coast	-0.020	-0.0097	-0.0081	-0.0028	-0.0046	-0.00034	-0.012	-0.0065
	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)
West Region	-0.088	-0.099	-0.022	-0.024	-0.023	-0.034	-0.065	-0.072
	(0.11)	(0.11)	(0.04)	(0.05)	(0.03)	(0.04)	(0.09)	(0.09)
East Region	0.16*	0.15*	-0.043	-0.051	-0.028	-0.034	0.21***	0.20***
	(0.08)	(0.08)	(0.04)	(0.05)	(0.03)	(0.03)	(0.06)	(0.06)
Constant	-0.61	0.71	5.13***	5.87***	3.59***	4.23***	-6.18***	-5.94***
	(2.04)	(2.25)	(1.40)	(1.58)	(0.83)	(0.92)	(1.41)	(1.58)
First stage F	236	161	236	161	236	161	5.14	4.25

Table 7: Reduced form and Model Impacts of Downgrading Expressways
Counterfactual-Actual Means with Standard Deviations in Parentheses

	Reduced Form			Model
	Highways become 25 kph	Port travel time at 25 kph	Both	Both
	(1)	(2)	(3)	(4)
Changes in population counts, no regional primate distinction	497,608	-516,872	0	0
	(414,346)	(404,330)	(381,028)	(345,336)
Changes in population counts with regional primate heterogeneity	508,379	-547,987	0	
	(577,749)	(401,093)	(533,654)	
Component: Changes in population in regional primate prefectures	-660,645	-656,733	-1,091,474	5,358
	(614,332)	(587,348)	(853,214)	(696,192)
Component: Changes in population in other prefectures	627,108	-536,942	110,853	-2,593
	(421,111)	(377,000)	(329,620)	(291,368)

Notes: Counterfactuals in Columns 1 and 2 are not normalized to sum to 0 change. Counterfactual in Column 3 is renormalized to sum to 0 aggregate population change. Model base counterfactual in column 4 is constructed to have zero aggregate population change.

Table 8: Market Potential Regressions

	log Prefecture GDP, 2010		log Prefecture Pop, 2010		log Prefecture GDP, 2010	
log GDP within 6 hour drive, 2010	-0.021	-0.10	-0.078*	-0.13**	0.071	0.044
	(0.09)	(0.12)	(0.04)	(0.06)	(0.07)	(0.09)
X Rank 1 Prefecture		0.15*		0.099**		0.038
		(0.09)		(0.04)		(0.06)
log Driving time to nearest	-0.16**	-0.19**	-0.10*	-0.13*	-0.043	-0.050
international port	(0.07)	(0.08)	(0.05)	(0.07)	(0.03)	(0.03)
X Rank 1 Prefecture		0.12		0.066		0.043
		(0.10)		(0.06)		(0.06)
log Prefecture Population, 2010					1.09***	1.13***
					(0.14)	(0.12)
Rank 1 Prefecture		-1.99	0.046	-1.19*		-0.64
		(1.23)	(0.04)	(0.65)		(0.82)
log Prefecture Area, 2005	-0.052	-0.083	-0.095**	-0.11**	0.050	0.042
	(0.09)	(0.10)	(0.04)	(0.05)	(0.07)	(0.08)
log Central City Area, 1990	-0.10**	-0.11*	-0.037	-0.041	-0.062	-0.060
	(0.05)	(0.05)	(0.03)	(0.03)	(0.04)	(0.04)
log Central City Population,	0.12**	0.12**	0.036	0.037	0.077	0.075
1982	(0.06)	(0.06)	(0.03)	(0.03)	(0.05)	(0.05)
log Central City Roughness	-0.049	-0.052	0.0059	0.0037	-0.054**	-0.056**
	(0.03)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)
log Prefecture Roughness	-0.023	-0.030	-0.0011	-0.0059	-0.022	-0.024
	(0.03)	(0.03)	(0.01)	(0.01)	(0.02)	(0.02)
Provincial Capital	0.64***	0.68***	0.34***	0.36***	0.27***	0.27***
	(0.10)	(0.11)	(0.05)	(0.06)	(0.09)	(0.09)
log Prefecture Population,	0.57***	0.58***	0.86***	0.87***	-0.38***	-0.39***
1982	(0.11)	(0.11)	(0.06)	(0.05)	(0.13)	(0.13)
Share Prefecture Population	0.50	0.57	-0.22	-0.22	0.72	0.82
with High School, 1982	(0.91)	(0.95)	(0.41)	(0.45)	(0.68)	(0.68)
Share Prefecture Population	1.98***	1.94***	-0.42	-0.45	2.41***	2.44***
in Manufacturing, 1982	(0.54)	(0.56)	(0.35)	(0.36)	(0.35)	(0.35)
log km to Coast	-0.021	-0.0077	-0.010	-0.0018	-0.010	-0.0056
	(0.03)	(0.04)	(0.01)	(0.02)	(0.03)	(0.03)
West Region	-0.10	-0.18	-0.063	-0.12	-0.021	-0.049
	(0.14)	(0.17)	(0.06)	(0.08)	(0.11)	(0.13)
East Region	0.17**	0.18**	-0.012	-0.0079	0.18***	0.18***
	(0.08)	(0.09)	(0.04)	(0.04)	(0.07)	(0.07)
Constant	-0.78	0.31	4.48***	5.21***	-5.61***	-5.58***
	(1.54)	(1.81)	(1.20)	(1.37)	(1.01)	(1.19)
First stage F	18.8	6.67	18.1	6.67	12.9	5.56

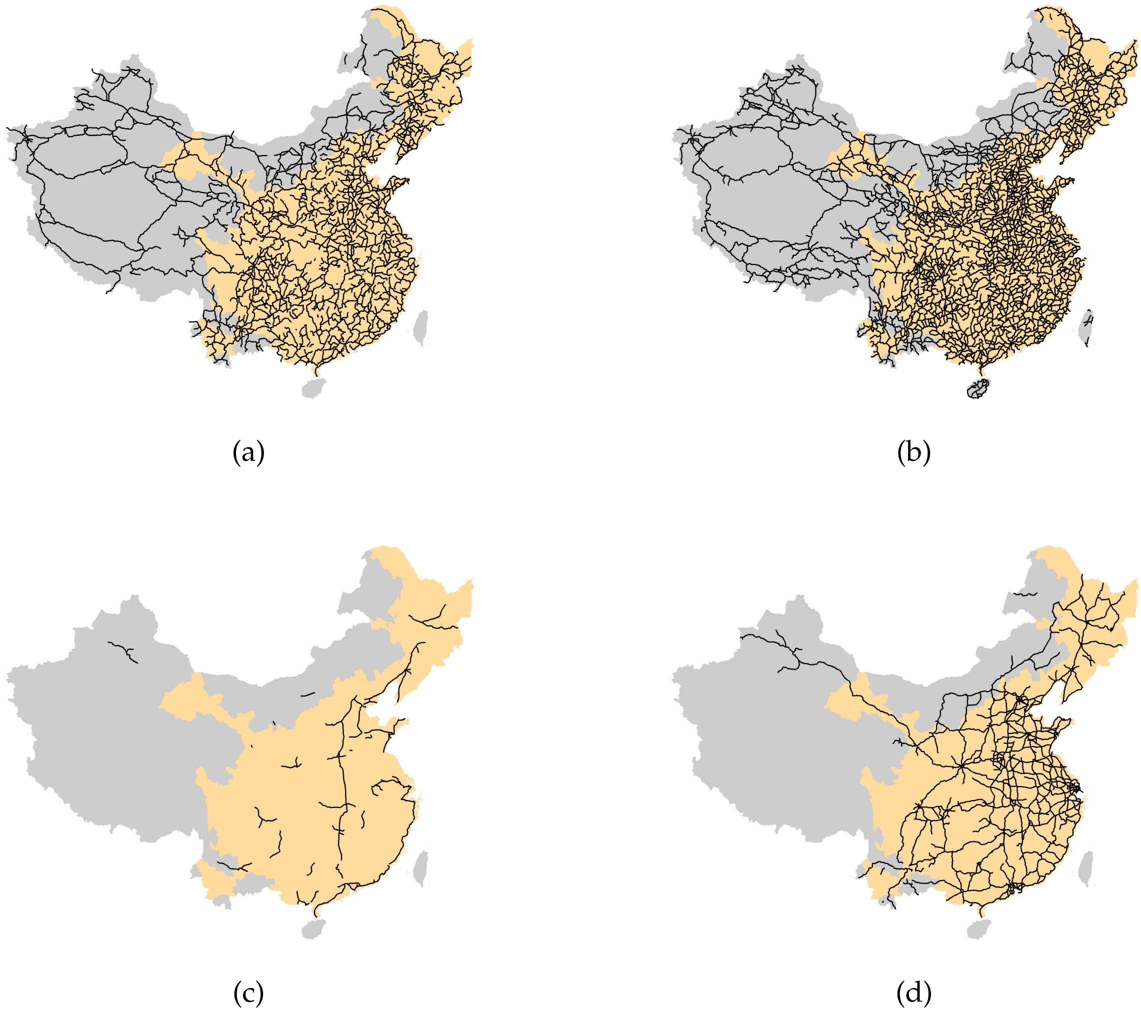
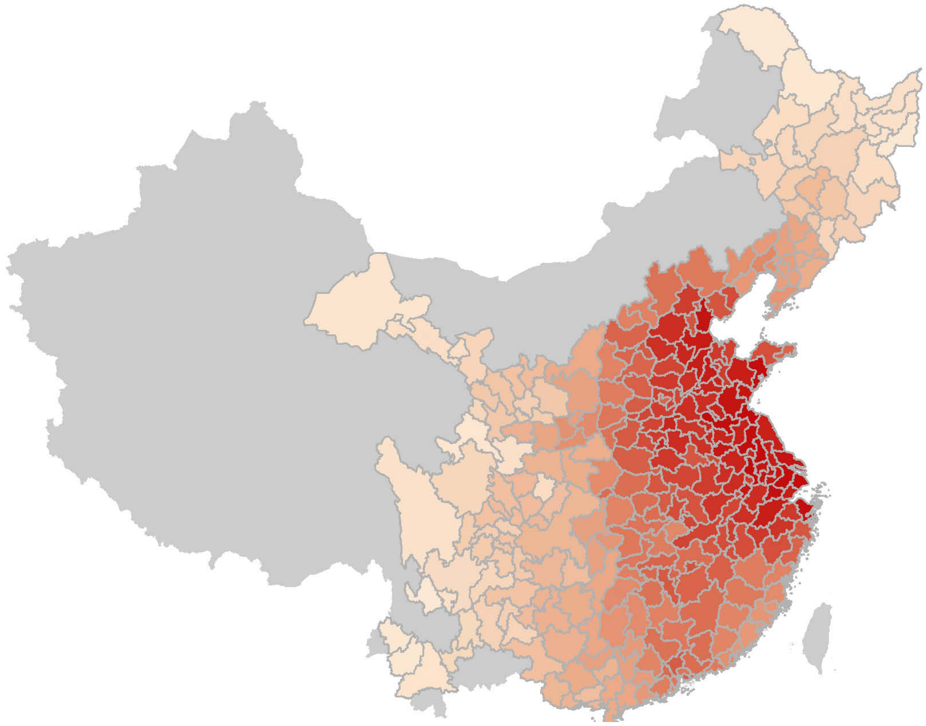
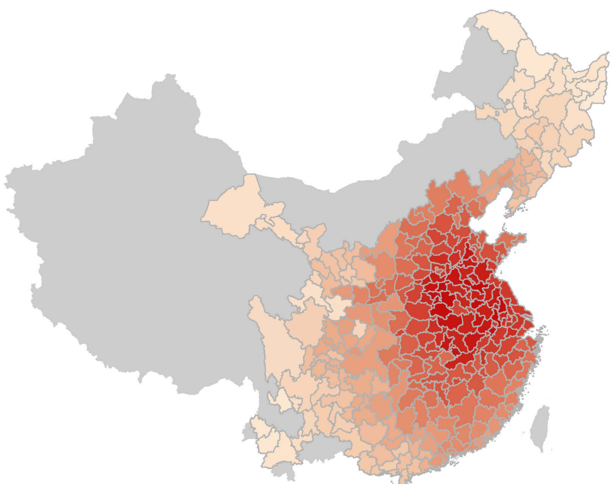


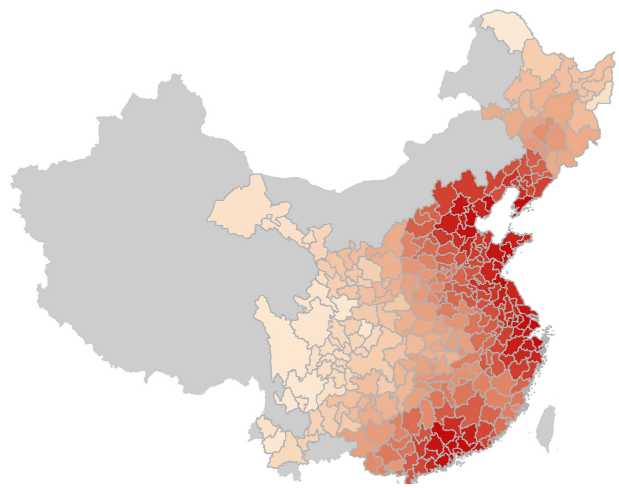
Figure 1: Illustration of Chinese Road and Highway networks: (a) 1962 national roads; (b) 1990 national roads; (c) 1999 limited access highways; (d) 2010 limited access highways. In all figures, the extent of our study area is indicated in pink.



(a)



(b)



(c)

Figure 2: Top panel (a) shows market access calculated from realized GDP and the observed transportation network. Colors indicate ordinal rank of the prefecture's market access, with darker colors indicating prefectures with larger market access values. Panel (b) shows the corresponding graph for the portion of market access determined by the domestic trade costs and GDP. Panel (c) is the corresponding graph for the export portion of market access.

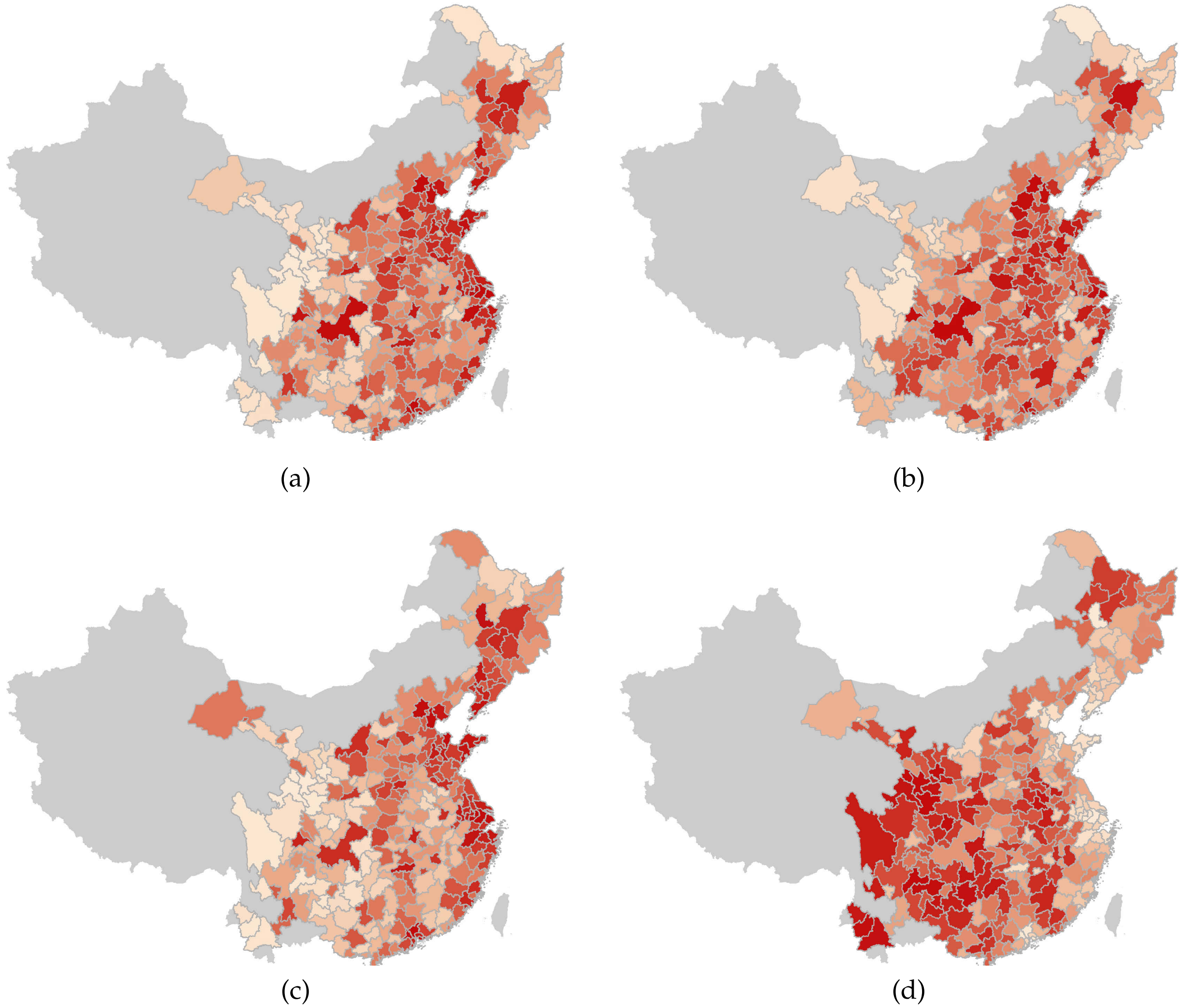
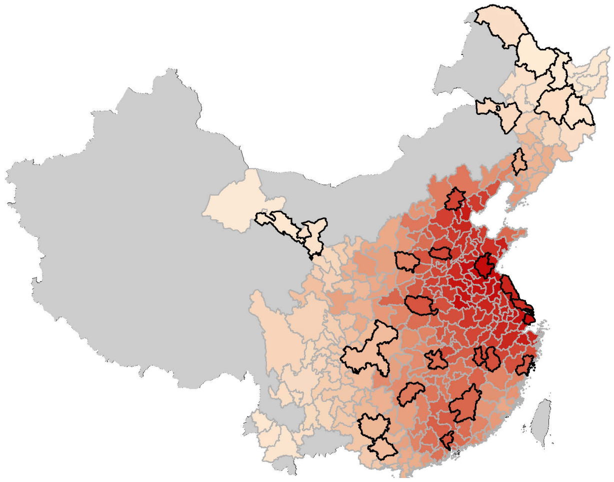
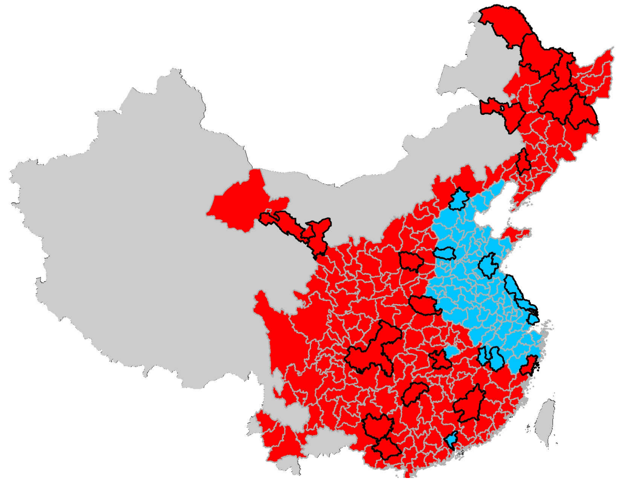


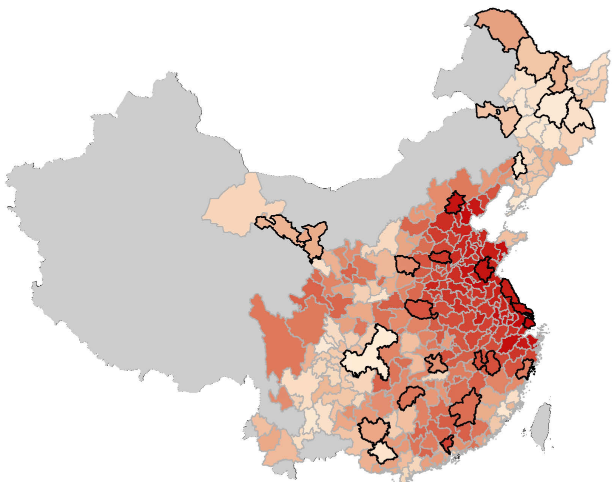
Figure 3: All panels illustrate rankings of prefectures, with darker colors indicating larger values of the relevant value: (a) observed 2010 GDP; (b) observed 2010 population; (c) estimated TFP, the model parameter ϵ; and (d) estimated amenity value, the model paramter A. Note that panels (c) and (d) show generally larger TFP near the coast and larger amenities in the West.



(a)

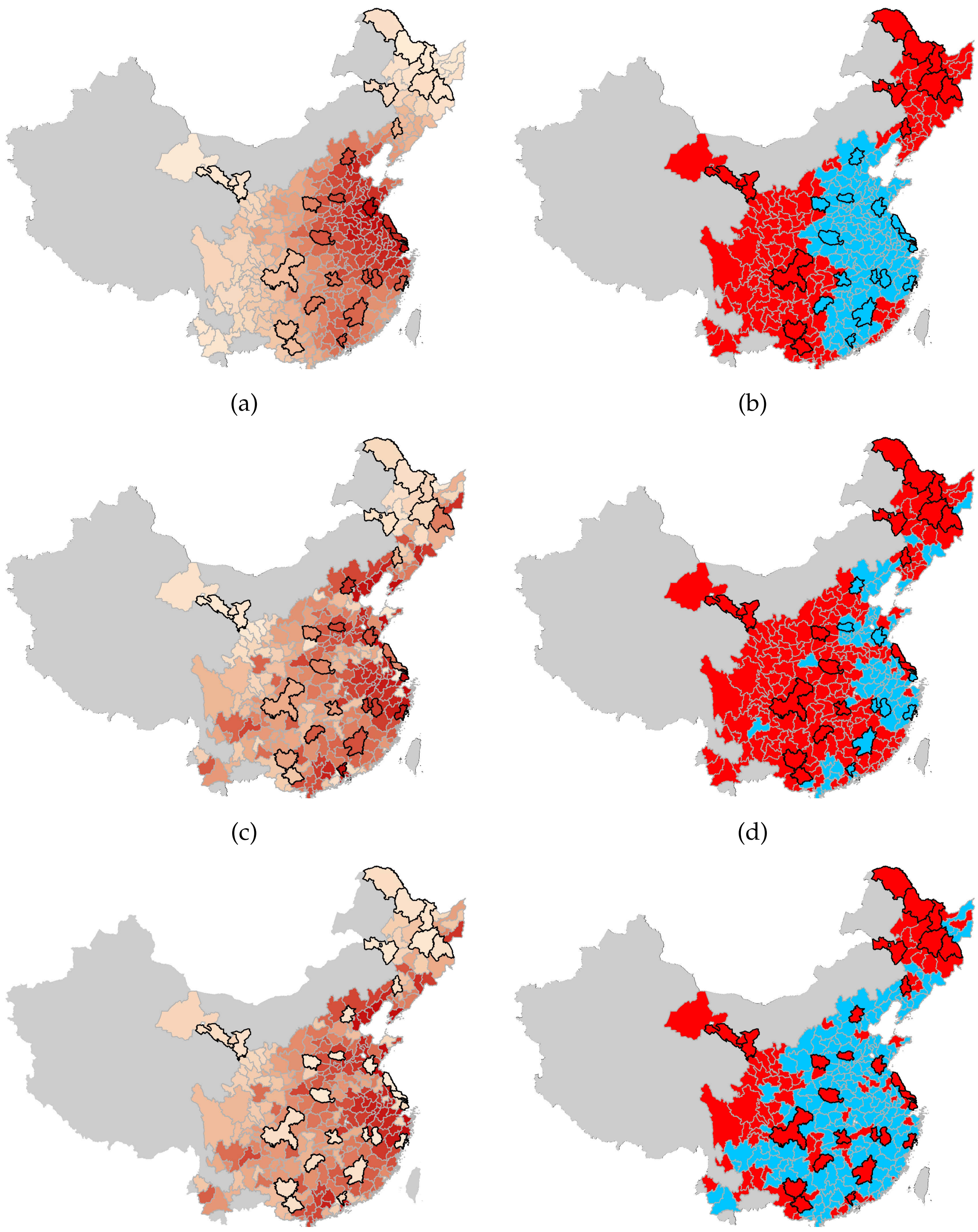


(b)



(c)

Figure 4: Counterfactual changes in GDP. Top is logs and bottom is levels. In the left column, colors indicate a prefectures ranking, darker colors indicate a larger increase in GDP under the counterfactual transportation network. In the right column, red indicates losers and blue indicates gainers. In all panels, highlighted prefectures are 'rank 1' as defined in the text.



(a) (b) (c) (d) (e) (f)

Figure 5: Counterfactual changes in logs of population. In the left column, colors indicate a prefectures ranking, darker colors indicate a larger increase in population under the counterfactual transportation network. In the right column, red indicates losers and blue indicates gainers. In all panels, highlighted prefectures are 'rank 1' as defined in the text. The top row indicates population changes predicted by the model. The second row indicates population changes under reduced form counterfactual 1. The third row indicates population changes under reduced form counterfactual 2.