Working paper

Highways, market access, and urban growth in China

Nathaniel Baum-Snow, Loren Brandt, Vernon Henderson, Matthew Turner, and Qinghua Zhang

January 2016

When citing this paper, please use the title and the following reference number:
C-89114-CHN-1

DIRECTED BY

FUNDED BY

Highways, Market Access, and Urban Growth in China*

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

January 20, 2016

Abstract

We investigate the effects of the construction of the national highway system in China on local economic outcomes. The analysis employs three main approaches. The first is based on a structural model of Ricardian trade that provides an explicit description of the general equilibrium effects of changes in the highway network. The second involves reduced form estimates of the casual effects highways, which accommodates the non-random assignment of highways across locations. The third approach is a hybrid of the first two. Technique matters. The structural model suggests that access to domestic markets, but not to export markets, increases economic output. The reduced form estimates suggest the opposite conclusion and also point to the importance of highways in the rise of regional primate cities. These reduced form findings are consistent with export driven growth policies and central or provincial government policies favoring regional primate cities. In addition to informing policy, our results raise concerns about the use of quantitative results from Ricardian trade models in isolation for understanding how and the extent to which infrastructure drives regional growth.

*The authors gratefully acknowledge financial support from the IGC (grant #1-VRU-VNEC-VXXXX-89114) and from the Global Research Program on Spatial Development of Cities, funded by the Multi Donor Trust Fund on Sustainable Urbanization of the World Bank and supported by the UK Department for International Development. We received helpful comments from Gerald Carlino, Edward Glaeser, Samuel Marden, Daniel Sturm and Junfu Zhang and many seminar participants. We thank Ying Chen for research assistance.

1 Introduction

Between 1990 and 2010, China constructed an extensive modern road network including a national system of limited access highways. We investigate the effects of this network on GDP, population and GDP per capita in Chinese prefectures. Our investigation faces two challenges. First, highways are not randomly assigned to locations within China and may have been allocated according to locations' productivity or attractiveness as a place to live. Second, output in each region depends on output in other regions through trade linkages. As a result, highway construction may have general equilibrium effects as trade and migration adjustments cause the impacts of highway construction near one prefecture to ripple through the country. This limits our ability to assign regions to treatment and control groups, a foundational requirement of well identified econometric analyses. This lack of a clear control group makes it difficult to distinguish highways' effects on aggregate growth from their effect on the distribution of economic activity across regions. Because there is no control group for 'all of China', identifying the effects of the new highway system on national outcomes requires strong structural assumptions. Consequently, despite their reliance on strong structural assumptions, economic geography models are indispensable for analyzing the effects of these roads on national aggregates and for evaluating welfare consequences.

To estimate causal effects of highways on prefecture output and population, we implement three distinct, parallel, research designs. As in Donaldson and Hornbeck (2015), Alder (2015) and Tombe and Zhu (2015), the first research design calibrates a general equilibrium model of Ricardian trade in the spirit of Eaton-Kortum (2002) [henceforth 'EK']. We generalize standard extant versions of the EK model to separately describe international and domestic market access effects. We calibrate our model and use it to conduct counterfactual exercises. In contrast to conventional econometric analyses, this approach accounts explicitly for general equilibrium effects. However, the price of an explicit model of general equilibrium effects is high. The model requires assumptions restricting transportation costs to affect population and output only through the one Ricardian channel and does not allow for the non-random assignment of network connections.

Using the model, we calculate output, population and welfare associated with various counterfactual road networks. In our main counterfactual example, we reduce expressway speeds in 2010 from 90 kph to the same 25 kph we assume for other roads. For a wide range of parameter values describing input cost shares and productivity dispersion, we 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. Associated reduced access to international markets has only tiny effects on output and welfare. We also find that the model predicts population gains for prefectures in the denser coastal area and losses for prefectures in the more sparsely populated interior.

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 a prefecture's main city and on travel time to the most accessible port. The reduced form research 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 validity of our instruments is easier to defend in the Chinese context. The price of resolving this endogeneity problem is, of course, that it is difficult to use this regression based research design to account for general equilibrium effects. The regression framework estimates slope coefficients, which should be interpreted as the relative gains or losses to one city resulting from a marginal change in its regional highway allocation relative to other locations.

In a variant of this exercise, following Head and Mayer (2004), Redding and Venables (2004) and Hanson (2005), we also 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 that the relationships between market potential, GDP and population are qualitatively similar to those of our raw infrastructure measures.

Reduced form estimates 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. This is a consequence of heterogeneity in the effects of highways on prefectures. Regional highways promote concentration of both output and population into regional primate prefectures, at the expense of other prefectures. Since there are more small prefectures than large, the average is negative. We speculate that the effect of roads on primate cities reflects migration and capital market policies that are not part of the Ricardian framework. We note that while primate prefectures are larger than average, our primate city definition does not simply pick out the largest cities. That is, the primate city effect cannot be solely attributed to size.

While regression results indicate that improvements in domestic market integration do not increase average output or population, they also indicate that improved access to international ports promotes growth in GDP, population and GDP per capita for all prefectures. 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 prefectures. This suggests that, in the context of export driven investment and growth policies (Branstetter and Lardy, 2008), better access to international markets has had a high return.

To summarize, counterfactual results from the EK model indicate that the highway expansion affected prefecture GDP and population by reducing the costs of domestic trade and that access to export markets was unimportant. Regression results suggest the opposite. Counterfactual results from the structural model suggest that the highway network led to decreased concentration of population and output near the coast, while the regression results suggest that highways played a role in the rise of regional primate cities but did not lead to a concentration of population near the coast. Technique matters. The choice between a structural EK and a regression based research design is fundamental to our understanding of the effects China's highway construction. To inform this choice, we provide three pieces of evidence, all of which favor the conclusion that the standard EK model that has been widely adopted to analyze related questions misses some quantitatively important features that are evident in the data.

First, the model systematically describes only about 5% of the variation in the GDP data. Our EK model can be sensibly partitioned into a systematic part describing trade and its implications and what are effectively prefecture 'fixed-effects', which capture locational fundamentals. We find that a regression of observed prefecture output on these prefectural fixed-effects gives an R^2 of about 0.95, leaving only 5% left to be explained by the systematic part of the model. This is low relative to commonly conducted city level regressions and means that model fails to incorporate important patterns in the data.

Second, we implement our third research design, which is hybrid of the first two. This research design follows Donaldson and Hornbeck (2015) and Alder (2015), and involves using the measures of market access implied by our implementation of the EK model as regressors in a conventional econometric analysis. Our calibration exercise requires that we assume values for all of the structural parameters in the model. We can use these same parameters to calculate the implied coefficients for this regression. Comparing estimated and calculated regression coefficients provides an informal over-identification test. The calculated regression coefficient values are qualitatively different from those we estimate.

This suggests that the structural model is mis-specified.

Third, we compare model generated with regression generated counterfactual prefecture populations. We begin with our regressions of prefectural population on roads within 450km. We adjust the roads measure to match that used in the main model counterfactual exercise described above. We assume that the counterfactual change in the road network leaves national population unchanged and compare the resulting regression based counterfactual to the corresponding EK counterfactual. The two counterfactual scenarios are quite different. The raw correlation between the two sets of population changes is less than 0.4 and these differences are systematic. Regression based counterfactuals indicate that removing the modern highway network would lead to population movements away from regional primate cities. On the other hand, model based counterfactuals indicate a concentration of population in the central coastal region when the modern highway network is removed.

We now summarize the case against relying exclusively on model based results for policy evaluation. First, the structural model does not allow for the possibility that roads are assigned endogenously. In principle the model could be amended to allow this, but this would require a non-trivial technical advance with additional structural assumptions. Second, the structural model has poor explanatory power. Third, the structural model fails the informal over-identification test suggested by our third, hybrid, research design. Fourth, the effects of counterfactual experiments in the model are at odds with the patterns that our regressions reveal in the data.

Of course, our regression based results are subject to one important critique: they do not provide a way to capture general equilibrium effects. This is a fundamental problem, but may not preclude using them to understand certain consequences of road construction. In particular, we think that the road network caused only tiny changes in the total population of China. Thus, for our investigation of the effects of roads on population relocation, the problem of determining the overall level of change does not arise. However, general equilibrium effects are central when it comes to accounting for the contributions of roads to overall economic growth.[1]

In all, the balance of evidence suggests that it is reasonable to rely on regression based

[1] Hillberry and Hummels (2003) find that domestic shipments overwhelmingly travel short distances. This suggests that the general equilibrium effects of regional road network improvements may not travel far. In this case, comparing prefectures with (quasi-random) roads to those without allows for recovery of the effects of the road network on output.

results to understand the impact of highway construction on population relocation, but that neither the structural nor the regression results is likely to give us a good understanding of how China's new road network impacted the overall level of economic activity. Clearly, this conclusion points to the need for a new generation of structural models that better describe what we observe and incorporate additional causal channels.

Our work relates to the literature in a number of ways. There is an active literature which adapts the EK model to study issues related to regional growth. For example, Alder (2015) uses the EK model to evaluate a new highway network in India; Donaldson and Hornbeck (2015) use the EK framework to evaluate the effects of the railroads on economic development in 19th the century US; Sotelo (2015) uses the EK model to evaluate a new system of paved roads in Peru; finally, Tombe and Zhu (2015) also apply the EK model to the study of regional growth in China. Given that each of these exercises relies on an EK framework similar to ours, and indeed, our model follows Donaldson and Hornbeck (2015) closely, it is natural to be concerned that some of the problems with the EK model that arise in our analysis may also be present in these other contexts.

With this said, the Tombe and Zhu (2015) analysis of China is of particular interest. The Tombe and Zhu implementation of the EK model is, arguably, superior to our own. Our model calibration is based on prefecture level measures of output and population, and on calculated transportation costs. Their analysis is based directly on inter-provincial trade flows. They also consider richer migration and production environments. In spite of this, their model leads them to a conclusion qualitatively similar to ours: domestic market integration is important and access to export markets is not. One obvious reason that our implementation of the EK model may fail to describe our data is simply that we have implemented it badly. That Tombe and Zhu investigate the implications of Ricardian trade using different data in a model whose details differ in many small ways from ours, yet still arrive at the same basic conclusion as we do suggests that the divergence between our different research designs is not a special case.

Importantly, the literature provides examples of structural models derived from alternative foundations that can be applied to the study of regional growth in general and to the evaluation of changes in transportation costs in particular. 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 and Khandelwal (2011) provide evidence that lower trade costs has fostered innovation through competition in India. Lower cost access to intermediate inputs (Fujita, Krugman and Venables, 1999) and innovative ideas (Alvarez, Buera and Lucas, 2013; Buera and Oberfeld, 2014) are additional mechanisms through which trade may promote growth. In light of our results, these frameworks clearly deserve further attention.

The theoretical literature describing central place theory is of particular interest. Central place theory orginates with Christaller (1933) and consists primarily of the conjecture that in any given region there should be a dominant city, the 'central place', that produces a full range of goods for sale to smaller more specialized cities, which in turn may produce goods for still smaller cities. This conjecture forms the basis for a theoretical literature that attempts to rationalize this geography from formal foundations. Krugman (1992) arguably provides such foundations in a geography consisting of two discrete locations. Fujita, Krugman and Mori (1999) and Tabuchi and Thisse (2011) develop specific general equilibrium models of such urban hierarchies along a line and around a circle. However, this literature has so far failed to produce a model which predicts the central place type geography in an empirically useful geography, i.e., a plane. In related empirical work, Glaeser and Ades (2005) investigate the determinants of urban primacy using cross-country data. To our knowledge, we are the first to provide econometric evidence for an 'urban hierarchy' at the regional level.

Apart from methodological issues, our estimates of the effects of reduced transport costs to ports echoes recent literature finding that improved access to ports fosters local economic growth in developing country contexts. Donaldson (2014), Banerjee, Duflo, and Qian (2012) and Storeygard (2016) 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 and Straub (2015) find the opposite for some rural counties served by roads in China and Brazil respectively.

2 Context and Data

2.1 Chinese Geography and Highways

The Chinese context is especially well-suited for our investigation. 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. In 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, with less than 5 percent of freight ton-miles moved by road. Since then, China has constructed an extensive intercity highway network. 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 we generate most results. 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, but left other cities with lower degrees of connectivity.

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

Figure 1 shows the national road networks in 1962, 1990, 1999 and 2010. We use the 1962 network to construct instruments for 2010 roads and travel costs. These travel costs assume speeds of 25 kph on local highways and 90 kph on expressways. Moving forward in time, we see the national expressway system developing a little between 1990 and 1999, with most of the country becoming linked between 2000 and 2010. The yellow shaded region is our study area. While we observe and incorporate some data from outside of this area, we use the 286 prefectures in this area as our primary estimation sample.

Because prefecture population is an outcome variable, it is important to understand the history of interregional population mobility in China. Before 2000, with the exception of a few coastal cities, cities hosted few inter-province or even inter-prefecture migrants. Migration was limited by the hukou system, which regulated and restricted migration between prefectures and imposed penalties for un-licensed migration. These restrictions were lifted in stages starting in the late 1990s and by 2010 un-licensed migration was no longer illegal, although the hukou system continues to restrict migrants' access to formal housing markets, schools, health care, and social security (Chan, 2008), particularly in larger cities.

Despite restrictions, a great deal of migration has occurred since 1990. In 1990 China's population was about 29% urban, a share that rose to about 50% by 2010. The change in urbanization has four 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 to urban areas. We ignore within prefecture migration and use the prefecture as the unit of analysis. Of course, the big story is the enormous growth in real GDP per capita in China over 20 years.

To evaluate the welfare implications of counterfactual policies, we must make assumptions about whether the observed pattern of development is or is not a steady state, and whether location patterns reflect free or costly mobility. The basic facts do not make these choices easy. On the one hand, it is difficult to argue that a place changing as rapidly as China is in a static equilibrium. On the other, it is also hard to argue that the marginal value of changing location can be large in an environment in which there exists so much mobility. In fact, the EK model relies on a static equilibrium; data limitations, described below, essentially preclude much beyond a cross-sectional analysis. However, we are agnostic about whether location patterns reflect free or costly mobility. Our regressions do not require us to take a stand on this issue and we calibrate versions of our EK model in these two alternative migration environments.

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. In our regression analysis, we consider 282 prefectures in Han China (about half the land area of China), omitting minority areas for data and contextual reasons, three cities that coincide with their prefectures, and one island prefecture. In our structural analysis we include the four omitted 'outlier' prefectures to respect the logical integrity of our general equilibrium model. In either case, our study area contains more than 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 and 2000 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. 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 of the nine most important international ports.

We assume travel at 25 kph on regular roads and 90 kph on highways. In our main EK model based counterfactual, we downgrade the 2010 highway network to 25 kph travel speeds. For our regression based research design, we measure road capacity as log 'efficiency kilometers' of highways within the 450 km disk centered on each prefecture city. To calculate this measure, assign a weight of one to road kilometers and a weight of $\frac{90}{25}$ to limited access highway kilometers. This is an intuitive way to put more weight on more important roads and facilitates a regression based counterfactual calculation in which we recalculate efficiency kilometers weighting both types of road equally. This allows us to evaluate regression based counterfactual scenarios that are qualitatively similar to our main EK model counterfactual.

The paper maps on which our digital maps are based were printed by the same publisher, drawn using the same projection and have similar legends. To the extent that it is possible, our data describe consistent sets of roads over time. With this said, the growth and improvement of China's road network was so dramatic that roads that were important enough to merit inclusion on the 1990 map probably bear little resemblance to roads that meet this standard in 2010, even if both roads receive the same designation in the legend.

Thus, we are reluctant to exploit the time series variation in our measures of highways. It is this data limitation and spotty GDP information for 1990 that motivate our focus on cross-sectional research designs.

Table 1 reports statistics on transportation related variables used in the paper.

2.3 1962 Roads and Modern Highways

The econometric part of our investigation recovers causal effects of 2010 highways and measures of access to markets facilitated by these highways on contemporaneous prefecture outcomes. While we postpone discussion of our estimating equations to Section 4, credible empirical results require exogenous variation in these 2010 highways. There are a few 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 their quality, had lower costs to build 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.[2] 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 also control for roughness to proxy for agricultural productivity, and for distance to the coast. Central city

[2] 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 2 Column 1 shows the result of regressing the log of 2010 efficiency units of roads within 450 km of prefecture cities on its counterpart in 1962 excluding own prefecture roads, other instruments, and control variables.[3] 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 built nearby, perhaps because manufactures traditionally traveled primarily by rail. Prefectures in the West had less highway length nearby, as is expected given the smaller amount of economic development in these areas. Results are similar when using larger or smaller distance rings. However, we do not have statistical power to separate out exogenous variation in road efficiency units for multiple rings simultaneously.

Column 2 of Table 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

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

may be expected.

Columns 1 and 2 of table 2 show that our instruments are strong predictors of endogenous variables of interest conditional on appropriate controls. These results also show 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

In order to recover per-capita GDP effects, in some of our regressions we control for 2010 prefecture population. To respond to the potential endogeneity of prefecture population growth, we use a migration shock instrument, following Bartik (1991) and Card (2001). This instrument relies on historical migration pathways to predict 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 2 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.

Remaining columns in Table 2 report first stage regressions for additional endogenous variables of interest that are explained in more detail below.

2.5 Rank 1 Cities

Each prefecture contains a single prefecture city. These prefecture cities are almost always the largest in the prefecture and are the administrative seat of the prefecture. We define a prefecture as ‘rank 1’ if and only if in 1982 its prefecture city was the largest in the set of prefectural cities within a 6 hour drive over the 1962 road network at 90 kph. Figures 5 and 6 highlight rank 1 prefectures with heavy black borders.

The choice of a 6 hour threshold is intended to measure a day's drive and leads us to clasify about 9% of our prefectures as rank 1. We have experimented with alternative definitions, primarily by varying the size of the regions over which primates are defined. Reducing driving time to 4 or 5 hours leads to a larger class of primate cities, but does not qualitatively change our results.

Our regression based research design will indicate that rank 1 prefectures are affected differently by changes in the road network than are subordinate prefectures. Given this, it is useful to describe the differences between the two sets of prefectures.

Figure 2 compares primate and non-primate prefectures. In all four panels, the solid line describes non-primate prefectures while the dashed line describes primate prefectures. Figure 2a shows the size distribution of primate and non-primate prefectures. We see that places classified as primate have larger prefecture populations on average than those that are non-primate and all large prefectures are primate. However, many small prefectures are also primate; the rank 1 classification is not just identifying large cities. This means that regressions to distinguish the effect of roads on rank 1 cities are actually revealing something about how cities at different places in regional hierarchies respond to roads instead of just differences in how large and small cities respond to roads.

Figures 2b-d compare the extent to which primate and non-primate prefectures produce for domestic and international markets. The central government designates National Develoment Zones (NDZs) to encourage local development. By 2005, there were 1.73 such zones in in an average rank 1 prefecture and only 0.5 in other prefectures. Using data from the 2007 annual survey of industry,[4] Figures 2b-d show the county share of firm book value, employment and value added for firms producing for export. Surprisingly, these figures show that in counties containing NDZ's, export firms account for a smaller share of book value, employment and value added in primate than in non-primate prefectures. Figures 2b-d each show that NDZ counties in rank 1 prefectures are less likely to be engaged in production for export.

[4]This survey reports on all plants with sales of more than 5m RMB in 2007.

3 Structural Research Design

In this section, we first develop a standard model that allows us to quantify Ricardian gains from trade integration and that can be calibrated with our data for China in 2010. The model allows us to evaluate consequences of various counterfactual road networks. It also provides the theoretically based measures of access to markets that we use in the hybrid research design described in Section 5.

3.1 Setup and Calibration

Our implementation of the EK model follows Donaldson and Hornbeck (2015) closely. Our primary innovation is to explicitly model trade flows between China and the rest of the world, in addition to the domestic trade that is the model's main focus. Because our framework is very similar to Donaldson and Hornbeck's, we provide only an outline of the model derivation and refer the interested reader to a technical appendix, to Donaldson and Hornbeck (2015) or to Alder (2015) for more detail.

We consider a set of 287 locations. 286 of these are Chinese prefectures and one is the 'rest of the world'. We denote domestic origin locations with i subscripts, domestic destination locations with j subscripts, and the rest of the world with x subscripts. The population of each prefecture is N_i and the population of China is $\overline{N}$. There are measure one of differentiated goods denoted $x(k)$ for $k \in [0, 1]$, with prices $p_j(k)$. Consumers in each prefecture j solve

$$\max A_j \left(\int_0^1 x(k)^{\frac{\sigma-1}{\sigma}} dk \right)^{\frac{\sigma}{\sigma-1}} \text{ s.t. } \int_0^1 p_j(k)x(k)dk = y_j,$$

where y_j is the consumer's endogenously determined wage and A_j is a local amenity, which reflects the utility loss in percentage terms that the consumer will experience from leaving prefecture j.

Production within each prefecture is Cobb-Douglas over land L, labor N and capital K such that output in each location is $Y_i = z_i L_i^{\alpha} N_i^{\gamma} K_i^{1-\alpha-\gamma}$ for each variety. 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 and the dispersion parameter θ is common across locations. We assume that the goods market is perfectly competitive. This ensures that income in each location is the aggregate value of trade flows to all locations, net of shipping costs. Capital is elastically supplied to each location. We get the initial equilibrium value of Chinese exports E in 2010 from the national accounts.

Based on our reading of the historical and Chinese production function literature, for our calibration exercise we use values of $\alpha = 0.1$ and $\gamma = 0.7$. Following EK, we assume $\theta = 5$. As we will see, most of our calibration results are not sensitive to the exact choice of these parameter values.

Shipping costs are iceberg and the cost of shipping one unit of any variety between i and j is $\tau_{ij} \geq 1$ units of that variety. To calculate τ_{ij}, we use

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

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

To calculate τ_{ix}, we use

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

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

Under a 'free mobility' assumption where population shifts across locations to equalize utility at a constant national level U, the following system of $3\times 286+2$ equations describes the equilibrium,

$$MA_i = \sum_j \tau_{ij}^{-\theta} \frac{Y_j}{MA_j} + \tau_{ix}^{-\theta} \frac{E}{\sum_j \frac{Y_j}{MA_j} \tau_{jx}^{-\theta}} \qquad i = 1, ..., 286 \qquad (2)$$

$$E = \frac{Y_x}{MA_x} \sum_j \left(\frac{Y_j}{MA_j} \right) \tau_{jx}^{-\theta} \qquad (3)$$

$$\ln Y_i = \epsilon_i + \frac{\gamma\theta}{1+\theta\alpha} \ln A_i - \frac{\gamma\theta}{1+\theta\alpha} \ln U + \frac{1+\gamma}{1+\theta\alpha} \ln MA_i \qquad i = 1, ..., 286 \qquad (4)$$

$$U = A_i \frac{\gamma Y_i}{N_i} MA_i^{1/\theta} \qquad i = 1, ..., 286 \qquad (5)$$

$$\overline{N} = \sum_j N_j \qquad (6)$$

where

$$\epsilon_i \equiv \frac{1}{1+\theta\alpha} \ln(\kappa_1 T_i) + \frac{\alpha\theta}{1+\theta\alpha} \ln(L_i/\alpha)$$

for $i = 1, ..., 286$, and $\kappa_1 = [\Gamma(\frac{\theta+1-\sigma}{\theta})]^{-\theta/(1-\sigma)} r^{-(1-\alpha-\gamma)/\theta}$. Although ϵ_i consolidates a number of prefecture characteristics, it also includes T_i, the parameter that determines the level of the prefectures' productivity dispersion. Given this, we abuse language slightly and refer to ϵ as the 'productivity of prefecture i'.

For some purposes, it will be useful to replace (4) with the equilibrium relationship between population and market access. The resulting equation is

$$\ln N_i = \epsilon_i - (\frac{\gamma\theta}{1+\alpha\theta} + 1)(\ln A_i - \ln U) + (\frac{1+\gamma}{1+\alpha\theta} + \frac{1}{\theta}) \ln MA_i. \qquad (7)$$

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.

Equation (2) describes 'market access' at each location. This elegant formula captures three intuitive features of the relationship between trade, output and distance. First, market access is increasing in the income of potential trading partners. Second, it is decreasing in the cost of moving goods between trading partners. Third, market access is decreasing in the extent to which potential trading partners have better access to competing trading partners. We refer to the first term in (2) as 'domestic market access', the second term in (2) as 'external market access'. To avoid confusion, we sometimes refer to the sum of these components as 'total market access'.

Equation (3) describes the relationship between exports to the rest of the world, rest of the world income and domestic income and market access. To understand this equation, note that if we substitute this expression for E into (2), the rest of the world is treated symmetrically with the other 286 trading units. That is, the model treats the rest of the world as a large remote domestic unit. Equation (3) and the asymmetric treatment of the rest of the world in equation (2) are necessary because we only observe total exports E. Differentiating external trading partners would require calculating market access for each other region around the world, requiring considerable additional data and calibration assumptions to implement.

Equation (4) describes equilibrium GDP. Intuitively, GDP increases in productivity, land, the local amenity and market access. The remaining equations describe utility and the population constraint.

We observe prefecture GDP Y_i, population N_i, the total value of exports E, pairwise trade costs τ_{ij}, and export costs τ_{jx} in our data. We calibrate α, γ and θ from the literature and normalize observed equilibrium utility to $U = 1$. Substituting these variables and parameters into the system (2)-(6), we are left with a system of $3 \times 286 + 2$ equations in $3 \times 286 + 2$ unknowns. Solving numerically, we can thus recover equilibrium values of MA_i, ϵ_i, A_i, $\overline{N}$ and 'world real income' $\frac{Y_x}{MA_x}$. Mechanically, the solution process involves first solving the system of equations (2) and (4) for MA_i and Y_i and then using these values to help solve the rest of the system. Note that the amenity and productivity variables operate like prefecture 'fixed-effects'. They adjust so that the model perfectly explains the data. With equilibrium values of A_i, ϵ_i and $\frac{Y_x}{MA_x}$ in hand, together with α, γ and θ, and counterfactual trade costs, τ'_{ij} and τ'_{jx}, the system (2)-(6) is, again, a system of $3 \times 286 + 2$ equations in $3 \times 286 + 2$ unknowns. Solving, we can thus calculate counterfactual levels of GDP, Y'_i, population, N'_i, market access MA'_i, utility, U', and exports E'.

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, (2) and (5) continue to hold, but equilibrium output is instead given by

$$\begin{aligned} \ln Y_i \quad = \quad & \frac{1}{1+\gamma\theta+\alpha\theta}\ln(\kappa_1 T_i) - \frac{\alpha\theta}{1+\gamma\theta+\alpha\theta}\ln(\alpha/L_i) - \frac{\gamma\theta}{1+\gamma\theta+\alpha\theta}\ln\gamma \\ & + \frac{\gamma\theta}{1+\theta\gamma+\theta\alpha}\ln N_i + \frac{1}{1+\gamma\theta+\alpha\theta}\ln MA_i. \end{aligned} \tag{8}$$

When we evaluate counterfactual road provision, we evaluate effects with and without population mobility.[5]

Table 1 presents summary statistics about total MA and its components while figure 3 maps the geography of MA. All maps show prefectures ranked from highest to lowest by color intensity. The pattern is clear. Figure 3 shows that domestic MA is spread smoothly over the country, as should be expected given its recursive nature. External MA is noticeably concentrated along the coast, also as 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 results presented later.

Figures 4a and 4b present maps of observed 2010 GDP and population, again by rank. While there are coastal concentrations, there are large economic centers in the interior as well. Figures 4c and d map values of A_i and ϵ_i backed out from the model, again using rank-color assignments. We see that higher productivity (higher ϵ) cities are on the coast and in traditional and newer industrial centers. High A prefectures are disproportionately in the fringe areas of Han China. This suggests not that these are high amenity locations, but instead that migration costs are higher from these locations.

We next examine the extent to which observed variation in GDP and population across prefectures is explained by the systematic parts of the model versus these fixed effects. From (4), we see that the units of both ϵ and $\ln A$ are log income. The data imply the following regression relationships:

$$\begin{aligned} \ln Y_i &= 3.210 + \underset{(0.11)}{0.390}\epsilon_i + \eta_i \\ \ln Y_i &= 6.98 + \underset{(0.39)}{2.12}\ln A_i + \underset{(0.08)}{0.944}\epsilon_i + \widetilde{\eta}_i \end{aligned}$$

Unsurprisingly ϵ is a powerful positive predictor of output in both equations. Moreover, the R^2 of $\ln Y$ on ϵ alone is 0.77 and the addition of $\ln A$ in the second regression increases it to 0.95. The systematic part of the model thus predicts only about 5% of the variation in the data. That is, standard Ricardian forces explain about 5% of the total variation in

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

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 (7), and is a linear function of $\ln A$ and ϵ. The analogous regressions to those above are,

$$\begin{aligned}
\ln N_i &= 3.03 - \underset{(0.32)}{0.125} \ln A_i + \nu_i \\
\ln N_i &= 7.86 + \underset{(0.46)}{0.935}\epsilon_i + \underset{(0.09)}{3.08} \ln A_i + \widetilde{\nu}_i
\end{aligned}$$

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. Adding ϵ raises the R^2 to 0.87. This suggests that the labor share of output is essentially constant across our sample.

Before we turn to our counterfactual results from the model, a comment about our hybrid research design is in order. Our hybrid research design, developed in Section 5, consists of regressions of output or population on market access based on (4),(7) and (8). First, in order to estimate these equations, we must assume something about how the error ϵ_i is related to MA_i. OLS cannot be used because both ϵ_i and MA_i are functions of unobserved productivity T_i and poorly measure land endowment L_i, meaning that there exists a structural relationship between the error term and a regressor. Second, we note that the system of equations (2)-(6) implicitly assumes that τ_{ij} and τ_{ix} are exogenous to the model, and in particular, that they are not influenced by components of output, Y_i including productivity, T_i. Treating this sort of endogeneity in the context of the EK model appears quite difficult and is beyond the scope of this paper. Dealing with the endogeneity of transportation costs econometrically is, however, entirely feasible and is an important part of the reduced form and hybrid research designs.

3.2 Counterfactual Results from the Model

We consider two main counterfactual scenarios. In the first, we examine the effects of increasing travel time by 5% between all locations. In the second, we impose 25 kph speeds on all 2010 highways, reverting all new limited access highways to secondary roads. In both counterfactual scenarios, we are also interested in assessing the relative importance of domestic and external market access. To accomplish this, we consider a counterfactual environment in which travel costs are calculated as if travel to international ports and between domestic trading partners occurred on separate networks whose speed can be adjusted independently.

Table 3 Panel A reports utility, GDP and exports for both classes of counterfactuals considered given free mobility (left side) and no mobility (right side). 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. Exports actually increase a bit as most large sources of Chinese demand are near the coast, and international shipping costs become relatively cheaper with rises in domestic shipping costs. GDP declines by 1.5 percent and prices go up, generating greater utility loss, both becuase of higher shipping costs and the reallocation of production capacity to less productive locations. Restricting mobility changes these numbers very slightly, since marginal labor flows from less to more productive locations must reflect small gaps in marginal productivities.

The second row of Panel A shows 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, exports rising by about 8.35 percent and GDP falling by 1.5 percent. Once again, a combination of falling GDP and rising prices results in reduced welfare. The final row in panel A shows almost no effect of changing external trade costs on outcomes, with only small resulting reductions in exports.

These counterfactuals point out that welfare changes reflect not only changes in GDP, but also changes in prices. This is important, especially given that single equation regressions ignore price changes. In addition, the model suggests that changing access to the coast has small effects. This will contrast with what regression results suggest.

Table 3 Panel B reports counterfactual levels of utility given 25 kph travel on all highways and various alternative parameter combinations given free mobility. Removing the 2010 highway network consistently causes about a 5% reduction in real GDP (utility) 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 real GDP to be extremely sensitive to these parameters in order to see a big effect. The exception, in the last row, is changes to the scaling of trade costs ρ. To a rough approximation, doubling the scale factor doubles the welfare impact of this reduction in travel speeds. Although we do not report them here, we have also experimented with changing the cost of international trade by changing the factor of 1.15 that determines τ_{ix} in equation (1), which some may consider too low. Such changes affect the equilibrium almost entirely by affecting the implied magnitude of world income, Y_x/CMA_x. Their other effects on counterfactual equilibria are tiny.

Figure 5a shows percent changes and 5c shows level changes in GDP from reducing highway travel speeds to 25 kph, again by rank-color intensity. Figure 5b shows the winning (blue) and losing (red) prefectures from this reduction. 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. Borders of regional primate cities are outlined in black in Figure 5; it is evident that these cities exhibit no differential effects of this treatment from their nearby locations.

4 Reduced Form Research Design

We now turn to our second, regression based, research design. This is a conventional instrumental variables estimation strategy with prefectures as the unit of analysis. Our object with this research design is to recover estimates of causal effects of highway connections and trade integration on prefecture GDP, population and GDP per capita.

4.1 Econometric 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. We consider variants of this measure in robustness checks. The second measure, E_{it} denotes the travel time over the road network to the nearest international port. In the context of the EK model, L and E can be thought of as reduced form analogs 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 determine outcomes of interest. To resolve this inference problem, as we discuss in Section 2.4, we rely on their 1962 counterparts as instruments.

Thus, a general statement of our 'Infrastructure only' estimation problem is

$$\ln y_{it} = a + \beta \ln L_{it} + \psi E_{it} + X_i \delta + u_{it} \quad (9)$$

$$L_{it} = a_1 + \beta_1 \ln L_{i62} + \psi_1 E_{i62} + X_i \delta_1 + \eta^1_{it} \quad (10)$$

$$E_{it} = a_2 + \beta_2 \ln L_{i62} + \psi_2 E_{i62} + X_i \delta_2 + \eta^2_{it}. \quad (11)$$

In (9), y denotes prefecture GDP or population and X denotes controls. Section 2 discussed instrument validity and the rationale behind the choices of control variables. Note that we take care to use the same set of controls in both the reduced form research design presented here, and in the hybrid research design developed below. This facilitates the comparison of the two sets of results. Because both research designs explain GDP and population and rely on the same instrument, 1962 road length within 450km excluding own prefecture, potential concerns about the conditional exogeneity of instruments lead to the same choices of controls in both cases. The prefecture area control performs double duty. It describes a factor of production and accounts for the possibility that larger 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 suspect may be correlated with an instrument and with 2010 GDP or population.

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; alternatively, there could be positive GDP effects everywhere.

We also carry out a parallel analysis which imposes constant population by explicitly controlling (and instrumenting) for 2010 prefecture population. The results of these regressions allow us to isolate variation in GDP after netting out migration effects, although we still cannot isolate the 'level' effect on average GDP per capita of the highway intervention.

Before turning to results, it is useful to consider how estimated effects of road efficiency units may be expected to differ from those of market access in the EK model. In short, the model tells us that interpretation of β and ψ in (9) may be complicated. In the context of the model, the treatment effect of nearby roads is increasing in local output's share of market access and is also a function of how GDP or population in each prefecture throughout the country depends on roads. However, a more straightforward interpretation arguably exists for ψ. This intuition, together with the theoretical literature on central place theory, leads us to our investigation of rank 1 cities and to the consideration of 'market potential' in an extension of our reduced form design in Section 4.3.

4.2 Results

Table 4 reports coefficient estimates from (9), 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 had 1.2 percent smaller populations. These results are at odds with what we expected in an environment with free mobility; we consider possible explanations below. Note that, absent controls, the relationships between regional roads and both population and output are positive. That is, higher GDP and population regions had more roads in 1962 and in 2010, but these locations gained less population than otherwise would have been expected given their other characteristics.

Next, we find strong evidence that better port connections lead to greater local output and population. Results in Columns 1 and 3 indicate that 10% less travel time to an international port leads 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 remote prefectures are unlikely to be marginal producers for export.

As we discuss above, we expect these estimates to conceal substantial heterogeneity across prefectures. Raising travel speeds to locations with low demand should have smaller effects than raising speeds to high demand locations. To investigate this possibility, we examine how treatment effects vary as a function of the importance of a prefecture in the local hierarchy.

Results in table 4 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 with better connections to nearby areas have larger GDPs and populations. 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 arranging hukou changes from nearby prefectures is easier in some areas of the country. We do not find any evidence that regional primate cities benefit more from faster port connections.

Regional hierarchy models offer one potential explanation for why reduced transport costs may favor rank 1 prefectures. Low rank prefectures lose 'protection' in markets for some domestically produced goods, as falling transportation costs allow rank 1 cities to compete in these markets. However, central government policy may also be at work. While export oriented firms often have access to foreign capital, either as direct investment or through international capital markets, firms oriented toward domestically consumed goods are more reliant on government controlled capital markets. Policies favoring primate cities in domestic capital markets and in hukou flexibility would also yield the patterns seen in Figure 2 since primate prefectures are more specialized in production for domestic consumption.

One potential identification concern about these results is that 1962 highways are correlated with unobservables about cities that are serially correlated. To allay this concern, Columns 5-6 of table 4 show population results differenced between 1990 and 2010. They are almost identical to the levels results in Columns 3-4. Because we have incomplete and poorly measured GDP data for 1990, we do not present 1990-2010 differenced GDP results. Analogous 1990 population level regressions yield small and insignificant coefficients on 2010 transport measures.

Columns 7-8 of Table 4 present regression results analogous to those in Columns 1-2 with the addition of a control for 2010 prefecture population. Since this regression predicts changes in GDP holding population fixed, it is essentially a regression explaining per capita GDP. This 2010 population control is instrumented with predicted migration flows, as is explained in Section 2.4. 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 per-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.

Many OLS results, reported in Table A1, are comparable to the IV results. This means that most of the roads that contribute to our efficiency measures were built along 1962 rights of way.

4.3 Extension to Market Potential Measures

An obvious difficulty with focusing attention on the effects of highways is the likelihood that the effects of highways on GDP and population will depend upon the characteristics of the locations that the treatment highways connect. One of the great strengths of the EK model is that it provides such an elegant description of this heterogeneity through the market access measure. A number of alternative theoretical foundations give rise to similar types of relationships between a location's output and its 'market potential'. Among these are Redding and Venables' (2004), Hanson's (2005) and Head and Mayer's (2005) adaptations of Fujita, Krugman and Venables' (1999) "New Economic Geography" model.

In order to detect these sorts of heterogenous treatment effects without resorting to structural estimation, we consider reduced form measures of market potential. In particular, we consider aggregate output reachable within a 6 hour drive over the road network as our primary market potential measure, denoted

$$MP_i = \sum_{j \neq i} Y_j 1(\text{hours of travel time}_{ij} < 6) \ . \tag{12}$$

This measure of market potential starkly imposes cheap trade within six hours drive and prohibitively expensive trade beyond. This cutoff is broadly consistent with observation in the US, where the preponderance of manufactured goods are shipped less than this distance (Hillberry and Hummels, 2005). Of course, nonparametric measures of output within various travel time bands or a 'gravity' measure which sums distance discounted GDP may be more informative. We experimented with various such alternatives and arrived at qualitatively similar results. Limits to first stage power preclude including more than one market potential measure at a time. We report results on the basis of the particularly simple formulation of market potential described by equation (12) with the caveat that our data cannot distinguish the effects of this measure from those of its cousins.

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

$$\ln y_i = s + \lambda \ln MP_i + \phi E_i + X_i\mu + \nu_i. \tag{13}$$

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.

Substituting the relationship between market potential and output, equation (13), into the definition of output, equation (12), we have $MP_i = \sum_j e^{s+\lambda \ln MP_j + \phi E_j + X_j\mu} e^{\nu_j} 1(\text{hours of travel time}_{ij} < 6)$, we see that $\ln MP_i$ is correlated with v_i by construction. Under strong assumptions, there are established techniques to recover parameters of this spatial lag model (Kelejian and Prucha, 2010).[6] However, we would like to allow for flexibility in model specification that these standard methods do not permit.

To accomplish this, we use an exogenous component of $\ln MP_i$ as an instrument - the km of 1962 roads within 450 km of each prefecture city excluding the own prefecture. Results in Table 2 show that this is a strong predictor of market potential. We use the same control variables as in regressions investigating effects of road efficiency units within 450km with rationales for this choice of control variables the same. This setup allows for direct comparisons of results in Tables 4, 5 and 6, because they only differ in the explanatory variable used to measure infrastructure.

Table 5 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, just as in the raw infrastructure regressions in table 4. As with the infrastructure results, we find that rank matters for the

[6] Gibbons, Overman and Pattacchini (2015) discuss the pitfalls of using these methods.

effects of market potential but not for port access.

Remarkably, regression results for market potential tell the same story as do those for efficiency roads within 450km. Specifically, prefectures that are better connected to international markets experience GDP, population and GDP per capita growth. Prefectures with better connections to nearby areas did no better in terms of GDP and lost population, resulting in potentially small GDP per capita gains. In both cases, improvements to highways divert population from rural prefectures to nearby primate cities.

Unlike for infrastructure regressions, there are large differences between OLS and IV market potential results. OLS estimates, reported in Table A2, suffer from the standard upward bias that comes with OLS estimation of models with positive spatial lag coefficients. OLS coefficients on market potential overstate true coefficients in all specifications.

5 Hybrid Research Design

The research designs employed in the prior section do not provide a theoretically founded approach for handling heterogenous treatment effects. It is intuitive that effects of a new highway should depend on the characteristics of the places it connects, and the preceding section considers only ad hoc solutions to this problem. We saw in Section 3 that the EK model delivers an exact calculation of the way that a highway affects the distribution of population and output, as mediated through the mechanism of Ricardian trade. Moreover, the model leads to equations describing the relationship between output and market access, equation (4), and between population and market access, equation (7). It seems natural to regard these two equations as regression equations and to use them to shed light on the underlying structural model. This is essentially the exercise conducted in Donaldson and Hornbeck (2015), though many details and their empirical context differ. Our employment of this research design is a hybrid of the reduced form regression and model calibration research designs in that it is regression based and concerned with the non-random assignment of highways but is also organized around estimating the relationship between a quantity derived from the EK model and outcome variables of interest.

Before we turn to regression results, we discuss two issues related to this research design. First is the possibility of an informal over-identification test. The second is an enumeration of inference problems that arise in estimating (7) or (4).

First, in order to recover quantitative information from the EK model in Section 3, we imposed values for the model's structural parameters. Using these values, we calculate the coefficient of MA in equation (4) as $\frac{1+\gamma}{1+\theta\alpha} = 1.13$. Varying structural parameter values within the ranges we consider in Table 3 always leads to a calculated MA coefficient of well under 2. We calculate the corresponding coefficient in the exogenous population version (8) as $\frac{1}{1+\gamma\theta+\alpha\theta} \approx 0.2$. Finally, we note that in the no mobility variant of the model, described by (8), the market access trivially has 0 impact on prefecture population. While our setting is not well suited to conduct a formal over-identification test, we compare the results from corresponding regressions to these calculated values to see if they are close.

The same inference problems arise in estimating (4) as we saw in our reduced form research design. Estimating equation (4) involves regressing prefecture GDP on a constant, prefecture land area and market access. 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 similar to the problem that arose in our market potential regressions and is also akin to the difficulty one faces in estimating spatial lag models. In addition, and as we have discussed above, the road network and the resulting τ_{ij} and τ_{ix}, may not be exogenous. Prefectures with high levels of output and strong trading links may receive more highways than others.

Consistent with the discussion in Section 2, and with the reduced form research design above, we instrument for log domestic MA using the km of 1962 roads within 450 km of the prefecture's main city but outside 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 2 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.[7]

Table 6 Columns 1 and 2 report regression results for total MA and for the separate domestic and external components. Results in Column 1 indicate that prefectures with 10 percent greater joint domestic and international market access had about 29 percent greater GDP. The coefficient on domestic market access in Column 2 is -8.8, relative to

[7] We use transport components of market access without incorporating (for example) lagged GDP or employment in order to eliminate the possibility that unobserved serially correlated components of these quantities introduce endogeneity into the MA instruments.

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 $0.3\frac{1+\gamma}{1+\theta\alpha}$.

The estimated coefficient on MA in Column 1 is 2.91, more than double the calibrated value of 1.13. Regression results in Column 2 are qualitatively consistent with the results of our reduced form our research design. Access to international markets is estimated to contribute more to growth than access to domestic markets, once again contradicting the results of model based counterfactuals presented in Section 3. Quantitatively, estimated market access coefficients in Column 2 are clearly unstable with large standard errors. Some of this may have to do with specification. Reducing the set of controls increases the domestic market access coefficient enough to make it insignificantly different from 0. These coefficient estimates clearly indicate that the model is missing something centrally important about the data generating process.

We now turn to the estimation of (7). Results in table 6 Columns 3 and 4 show that total market access is not related to prefecture population and that prefectures with better access to external markets gain population while prefectures with better domestic market access lose population. We note that this second result holds only conditional on controls. Without controls the coefficient on domestic market access goes to 0. As with market potential results, analogous OLS estimates in Table A3 differ markedly.

To characterize equilibrium under an assumption of no population mobility, we estimate (8). This amounts to estimating the same regression equation as for (4) with the addition of a control for prefecture population. Columns 5 and 6 of Table 6 show these results. In these regressions, coefficients on 2010 log prefecture population (instrumented as explained above) are 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 value of 0.2 calculated from 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 in other prefectures. However, becoming better connected to external markets is likely to be welfare enhancing.

6 Reduced Form versus Model Counterfactuals

We have so far compared the results of model based counterfactuals to regression results from our reduced form or hybrid research designs. Since the model based counterfactuals describe general equilibrium changes while the other research designs consider local changes, all else equal, these objects are difficult to compare quantitatively.

To accomplish a more direct comparison, we impose a fixed national population constraint on regression based counterfactual population estimates. In particular, we consider the hypothetical reduction of highway speeds to 25 kph, calculate the implied population change for each prefecture, and then adjust each prefecture's population by a constant to equalize initial and final aggregate populations. Since aggregate GDP cannot be assumed constant under counterfactual road networks, we do not consider the corresponding exercise for GDP.

Table 7 shows the implied effects on prefecture populations. Columns 1-3 present reduced form counterfactual results calculated using the regression equation without regional primate city effects in Row 1 and with these effects in Rows 2-4. Row 2 shows average effects across cities while Rows 3 and 4 break out gains and losses for regional primate versus all other cities. Column 1 separately examines effects of giving expressways a weight of 1 rather than 90/25 in the efficiency units calculation. Column 2 separately examines effects of reducing highway travel speeds to ports to 25 kph. In these two columns we do not adjust to keep aggregate population constant because there is no clear way to do so. In Column 3 we show both operating together with all city populations adjusted by the same proportion so the net overall change is zero. Numbers in parentheses are the standard deviations of changes and show the degree of churning.

Consistent with Table 4's regression results, results in Table 7 indicate that cities gain population with reduced local access and lose population with reduced port access. In Rows 1 or 2, summing the average effects of the two components results in a small net loss. All types of cities suffer from reduced access to the coast, relative to those with better access. Regional primate cities suffer from reduced local market access, while other cities reclaim population from the regional primates under the counterfactual.

Changes under the EK model counterfactual are described in Column 4. Most stark are the results in Rows 3 and 4, indicating how the model predicts regional primates to exhibit small average gains while other cities have small average losses from transport reductions, unlike the regression results which generate very large changes that go in opposite directions. In addition, the amount of population churning implied by the model is somewhat below that predicted from the regressions in all rows. The final column reports correlation coefficients of reduced form population changes for individual cities as reported in column 3 with model changes in column 4. Overall, this correlation is 0.34, but this number is almost entirely driven by non-primate cities. For primate cities, the correlation is 0.03. Clearly, the model is missing something important about the urban hierarchy.

Figure 6 presents maps of model and reduced form population responses to reducing highway speeds to 25 kph. The top row shows results from the model, the second row shows results using the baseline reduced form regression equation, and the bottom row shows results from the primate city interacted regression equation. The left column shows percent changes using a rank-color scheme and the right column shows prefectures gaining population in blue and those losing population in red. Borders of regional primate cities are outlined in black.

As with GDP in Figure 5, winning prefectures as predicted by the model and reduced form analysis are primarily on or near the coast in the denser parts of the country. The most intense gainers are on the Beijing-Shanghai axis and their hinterlands. In general, the model generated population changes are much smoother over space then the reduced form ones. Even without distinguishing regional primate differentials, in Figures 6c and 6d we see smaller and less universal gains for the dense coastal areas predicted using estimated treatment effects than from the model. In addition, there are now interior gainers, who have lower domestic market access. Figures 6e and 6f show the role of regional primate cities, who are the intense losers from reduced regional networks. Contrasts between model predictions in 6a and 6b with the reduced form evidence in 6e and 6f are stark.

7 Conclusion

This paper analyzes impacts of the construction of the expressway network in China on the output and population of prefectures. Central to our study is the comparison of results from quantitative analysis of the workhorse EK model with treatment effect estimates that make use of historical roads as a source of exogenous variation. Evidence from the model indicates that domestic Ricardian trade forces are centrally important for driving prefecture population and GDP responses to changes in the transport network and welfare gains of about 5%. In contrast, reduced form estimates indicate that highways promote local growth through improved linkages to export nodes and concentrate economic activity toward regional primate cities. Regression specifications that match model structural equations deliver results that broadly echo the conclusions from the reduced form analysis, though their instability and inability to match calibrated parameters is evidence of model mis-specification.

In summary, our results indicate that important mechanisms must be at play that are not incorporated into the workhorse EK model. The model's almost exclusive focus on Ricardian forces limits its utility for evaluating effects of changing interregional transport costs, at least for China. The model's missing mechanisms may be related to Chinese government policies or be more fundamental. The Chinese government explicitly subsidizes exports and uses hukou and capital market policies to channel resources for domestic development to regional primate cities. More fundamentally, the EK model's arbitrary regional units, exogeneity of local productivity distributions, simple production structure and lack of accommodation of urban hierarchies may limit its utility in this case. Despite these potential limitations, some such model is essential for evaluating welfare consequences of transport policies. We hope that this study has established some directions in which to develop this class of models to generate implications that more closely align with treatment effects seen in the data.

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 and Kortum (2002) demonstrate that the equilibrium value of trade flows between each pair of domestic origin and destination locations is given by

$$X_{ij} = \kappa_1 T_i (q_i^a w_i^\gamma)^{-\theta} \tau_{ij}^{-\theta} \frac{Y_j}{CMA_j}. \tag{14}$$

In (14), 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 (14), 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:

$$Y_i = \kappa_1 T_i (q_i^a w_i^\gamma)^{-\theta} \left(\sum_j \tau_{ij}^{-\theta} \frac{Y_j}{CMA_j} + \tau_{ix}^{-\theta} \frac{E}{\sum_i \kappa_1 T_i (q_i^a w_i^\gamma)^{-\theta} \tau_{ix}^{-\theta}} \right) \tag{15}$$

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 (15) 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 (15), 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 determine how E under various counter-factual scenarios.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Donaldson, D. (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,' *Quarterly 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(4): 1407-1440.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Tombe T. and X. Zhu (2015) ‘Trade, Migration, and Productivity: A Quantitative Analysis of China,’ manuscript.

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

Table 1: Predictors and Instruments
Means and (Standard Deviations)

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)

Notes: The sample includes 282 prefectures in Han China, as is explained in the text. Market access variables are calculated as explained in the text.

Table 2: First Stage Regressions

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

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

Table 3: Counterfactual Results from the Quantitative Model
Means Across Prefectures Relative to Baseline of 1

Panel A: Counterfactual Results

	Free Mobility			No Mobility		
	Utility	Exports	GDP	Utility	Exports	GDP
Set All Highway Speeds to 25 kph	0.948	1.012	0.985	0.945	1.005	0.975
Increase all travel minutes by 5 percent	0.960	1.082	0.985	0.958	1.083	0.980
Increase domestic travel minutes by 5 percent	0.960	1.085	0.985	0.958	1.083	0.980
Increase travel minutes to port by 5 percent	1.000	0.998	1.000	1.000	0.997	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. Utility in the free mobility case is constant national real income. In the no moblity case, Utility is mean prefectural real income as a fraction of observed mean prefecture real income.

Table 4: 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 5: 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

Table 6: 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 7: Reduced form and Model Impacts of Downgrading Expressways
Counterfactual-Actual Means and (Standard Deviations)

	Reduced Form			**Model**	**RF-Model Correlation**
	Highways become 25 kph (1)	Port travel time at 25 kph (2)	Both (3)	Both (4)	Both (5)
Changes in population counts, no regional primate distinction	497,608	-516,872	0	0	0.34
	(414,346)	(404,330)	(381,028)	(345,336)	
Changes in population counts with regional primate heterogeneity	508,379	-547,987	0	0	0.22
	(577,749)	(401,093)	(533,654)	(345,336)	
Component: Changes in population in regional primate prefectures	-660,645	-656,733	-1,091,474	5,358	0.03
	(614,332)	(587,348)	(853,214)	(696,192)	
Component: Changes in population in other prefectures	627,108	-536,942	110,853	-2,593	0.42
	(421,111)	(377,000)	(329,620)	(291,368)	

Notes: Counterfactuals in Columns 1 and 2 are not normalized to sum to 0 change. Counterfactuals in Column 3 are normalized to sum to 0 aggregate population change. Model based counterfactuals in column 4 are constructed to have zero aggregate population change

Table A1: Infrastructure Regressions - OLS

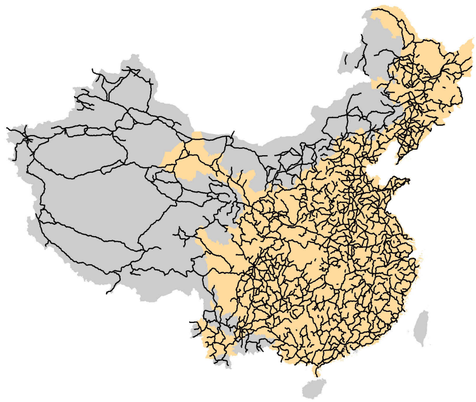
	log Prefecture GDP, 2010		log Prefecture Pop, 2010		D_censuspop9010_pref		log Prefecture GDP, 2010	
log Road Eff. Units within	0.087	0.027	-0.067	-0.095	-0.073*	-0.100**	0.17*	0.14
450 km of Prefecture City	(0.12)	(0.12)	(0.06)	(0.06)	(0.04)	(0.04)	(0.09)	(0.10)
X Rank 1 Prefecture		0.40*		0.24**		0.20**		0.11
		(0.22)		(0.11)		(0.08)		(0.18)
log Driving time to nearest	-0.14***	-0.16***	-0.098***	-0.11***	-0.068***	-0.072***	-0.017	-0.027
international port	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)	(0.03)	(0.03)
X Rank 1 Prefecture		0.096		0.043		0.018		0.043
		(0.06)		(0.03)		(0.02)		(0.05)
log Prefecture Population, 2010							1.21***	1.21***
							(0.10)	(0.10)
Rank 1 Prefecture		-4.74*		-2.74**		-2.24**		-1.43
		(2.51)		(1.26)		(0.92)		(2.02)
log Prefecture Area, 2005	-0.018	-0.033	-0.045	-0.055*	-0.038	-0.040*	0.036	0.033
	(0.06)	(0.07)	(0.03)	(0.03)	(0.02)	(0.02)	(0.05)	(0.05)
log Central City Area, 1990	-0.097**	-0.087**	-0.032	-0.028	-0.024*	-0.022	-0.058*	-0.053
	(0.04)	(0.04)	(0.02)	(0.02)	(0.01)	(0.01)	(0.03)	(0.03)
log Central City Population,	0.12**	0.11*	0.035	0.029	0.033	0.030	0.077*	0.072
1982	(0.06)	(0.06)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)	(0.04)
log Central City Roughness	-0.049	-0.051	0.0048	0.0049	0.0045	0.0031	-0.055**	-0.057**
	(0.04)	(0.04)	(0.02)	(0.02)	(0.01)	(0.01)	(0.03)	(0.03)
log Prefecture Roughness	-0.019	-0.025	0.0010	-0.0024	0.0039	0.0016	-0.021	-0.022
	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Provincial Capital	0.63***	0.66***	0.35***	0.36***	0.25***	0.26***	0.20*	0.22*
	(0.13)	(0.13)	(0.07)	(0.07)	(0.05)	(0.05)	(0.11)	(0.11)
log Prefecture Population,	0.53***	0.54***	0.81***	0.81***	-0.11***	-0.12***	-0.45***	-0.44***
1982	(0.07)	(0.08)	(0.04)	(0.04)	(0.03)	(0.03)	(0.10)	(0.10)
Share Prefecture Population	0.66	0.87	-0.19	-0.13	-0.31	-0.33	0.89	1.03
with High School, 1982	(0.88)	(0.91)	(0.45)	(0.46)	(0.33)	(0.33)	(0.70)	(0.72)
Share Prefecture Population	2.14***	2.15***	-0.43*	-0.43*	-0.042	-0.032	2.65***	2.66***
in Manufacturing, 1982	(0.49)	(0.49)	(0.25)	(0.24)	(0.18)	(0.18)	(0.39)	(0.39)
log km to Coast	-0.033	-0.026	-0.012	-0.0091	-0.0090	-0.0061	-0.018	-0.015
	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
West Region	-0.046	-0.037	-0.0017	0.0052	-0.0025	-0.0060	-0.044	-0.043
	(0.10)	(0.11)	(0.05)	(0.05)	(0.04)	(0.04)	(0.08)	(0.08)
East Region	0.18**	0.17*	-0.039	-0.042	-0.024	-0.027	0.23***	0.22***
	(0.09)	(0.09)	(0.04)	(0.04)	(0.03)	(0.03)	(0.07)	(0.07)
Constant	-1.89	-0.99	4.62***	5.17***	3.07***	3.55***	-7.48***	-7.22***
	(1.49)	(1.58)	(0.75)	(0.79)	(0.55)	(0.58)	(1.26)	(1.36)
R-Squared	0.77	0.78	0.88	0.88	0.42	0.44	0.86	0.86

Table A2: Market Potential Regressions - OLS

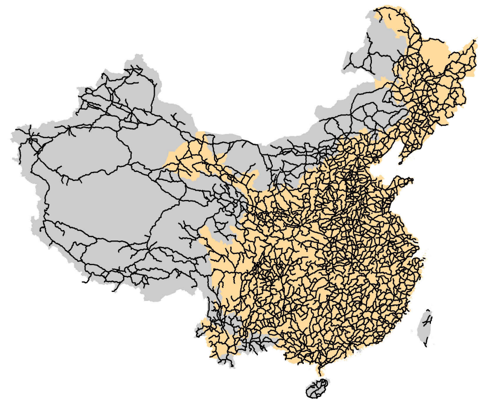
	log Prefecture GDP, 2010		log Prefecture Pop, 2010		log Prefecture GDP, 2010	
log GDP within 6 hour drive, 2010	0.12***	0.12***	-0.018	-0.025	0.15***	0.15***
	(0.04)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)
X Rank 1 Prefecture		0.0019		0.033		-0.039
		(0.05)		(0.03)		(0.04)
log Driving time to nearest	-0.12***	-0.13***	-0.096***	-0.10***	-0.0017	-0.0020
international port	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
X Rank 1 Prefecture		0.054		0.041		0.0033
		(0.07)		(0.03)		(0.05)
log Prefecture Population, 2010					1.23***	1.24***
					(0.09)	(0.09)
Rank 1 Prefecture		-0.29	0.065	-0.46		0.27
		(0.69)	(0.05)	(0.35)		(0.54)
log Prefecture Area, 2005	0.056	0.039	-0.051	-0.054	0.11**	0.11**
	(0.06)	(0.07)	(0.03)	(0.03)	(0.05)	(0.05)
log Central City Area, 1990	-0.091**	-0.087**	-0.033*	-0.032	-0.052*	-0.047
	(0.04)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)
log Central City Population,	0.12**	0.11**	0.037	0.036	0.072*	0.068
1982	(0.05)	(0.06)	(0.03)	(0.03)	(0.04)	(0.04)
log Central City Roughness	-0.049	-0.046	0.0070	0.0070	-0.055**	-0.055**
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
log Prefecture Roughness	-0.016	-0.018	0.0012	-0.00038	-0.019	-0.017
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Provincial Capital	0.64***	0.64***	0.33***	0.34***	0.22**	0.22**
	(0.13)	(0.13)	(0.07)	(0.07)	(0.11)	(0.11)
log Prefecture Population,	0.45***	0.47***	0.81***	0.81***	-0.55***	-0.53***
1982	(0.07)	(0.08)	(0.04)	(0.04)	(0.09)	(0.10)
Share Prefecture Population	0.72	0.90	-0.15	-0.067	0.88	0.99
with High School, 1982	(0.86)	(0.89)	(0.44)	(0.45)	(0.67)	(0.69)
Share Prefecture Population	2.13***	2.10***	-0.38	-0.39	2.57***	2.58***
in Manufacturing, 1982	(0.45)	(0.46)	(0.23)	(0.23)	(0.36)	(0.36)
log km to Coast	-0.040	-0.039	-0.016	-0.015	-0.020	-0.020
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
West Region	0.063	0.083	0.0062	0.0065	0.063	0.075
	(0.10)	(0.11)	(0.05)	(0.05)	(0.08)	(0.08)
East Region	0.14*	0.14*	-0.025	-0.026	0.18***	0.18***
	(0.09)	(0.09)	(0.04)	(0.04)	(0.07)	(0.07)
Constant	-1.81**	-1.78*	4.14***	4.29***	-6.79***	-7.09***
	(0.90)	(0.95)	(0.47)	(0.48)	(0.79)	(0.84)
R-Squared	0.78	0.78	0.88	0.88	0.87	0.87

Table A3: Market Access Regressions - OLS

	log Prefecture GDP, 2010		log Prefecture Pop, 2010		log Prefecture GDP, 2010	
log Market Access	3.60***		0.90		2.51***	
	(1.08)		(0.57)		(0.83)	
log Domestic Market Access		1.69		-0.97		2.89***
		(1.22)		(0.63)		(0.95)
log External Market Access		2.22*		2.41***		-0.77
		(1.27)		(0.65)		(1.00)
lcensuspop2010_pref					1.21***	1.24***
					(0.09)	(0.09)
log Prefecture Area, 2005	0.027	0.021	-0.026	-0.036	0.059	0.065
	(0.06)	(0.07)	(0.03)	(0.03)	(0.05)	(0.05)
log Central City Area, 1990	-0.081**	-0.083**	-0.022	-0.025	-0.054*	-0.052*
	(0.04)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)
log Central City Population,	0.13**	0.13**	0.035	0.035	0.084*	0.083*
1982	(0.06)	(0.06)	(0.03)	(0.03)	(0.04)	(0.04)
log Central City Roughness	-0.060*	-0.059*	-0.00083	-0.000094	-0.059**	-0.059**
	(0.04)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)
log Prefecture Roughness	-0.011	-0.010	0.0058	0.0077	-0.018	-0.020
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Provincial Capital	0.59***	0.60***	0.32***	0.34***	0.20*	0.18
	(0.13)	(0.13)	(0.07)	(0.07)	(0.11)	(0.11)
log Prefecture Population,	0.49***	0.50***	0.80***	0.81***	-0.48***	-0.51***
1982	(0.08)	(0.08)	(0.04)	(0.04)	(0.09)	(0.09)
Share Prefecture Population	1.07	0.94	0.15	-0.072	0.89	1.03
with High School, 1982	(0.89)	(0.90)	(0.47)	(0.46)	(0.68)	(0.69)
Share Prefecture Population	2.59***	2.48***	-0.039	-0.21	2.63***	2.74***
in Manufacturing, 1982	(0.47)	(0.47)	(0.24)	(0.25)	(0.36)	(0.37)
log km to Coast	-0.060**	-0.051*	-0.039***	-0.024*	-0.012	-0.021
	(0.02)	(0.03)	(0.01)	(0.01)	(0.02)	(0.02)
West Region	0.043	0.097	0.041	0.13**	-0.0074	-0.065
	(0.11)	(0.12)	(0.06)	(0.06)	(0.08)	(0.09)
East Region	0.22**	0.17*	0.018	-0.070	0.20***	0.26***
	(0.09)	(0.10)	(0.05)	(0.05)	(0.07)	(0.08)
Constant	-25.3***	-23.8***	-2.57	-3.15	-22.1***	-19.9***
	(7.00)	(6.43)	(3.68)	(3.32)	(5.40)	(4.96)
R-Squared	0.76	0.76	0.87	0.87	0.86	0.86



(a)



(b)

(c)

(d)

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.

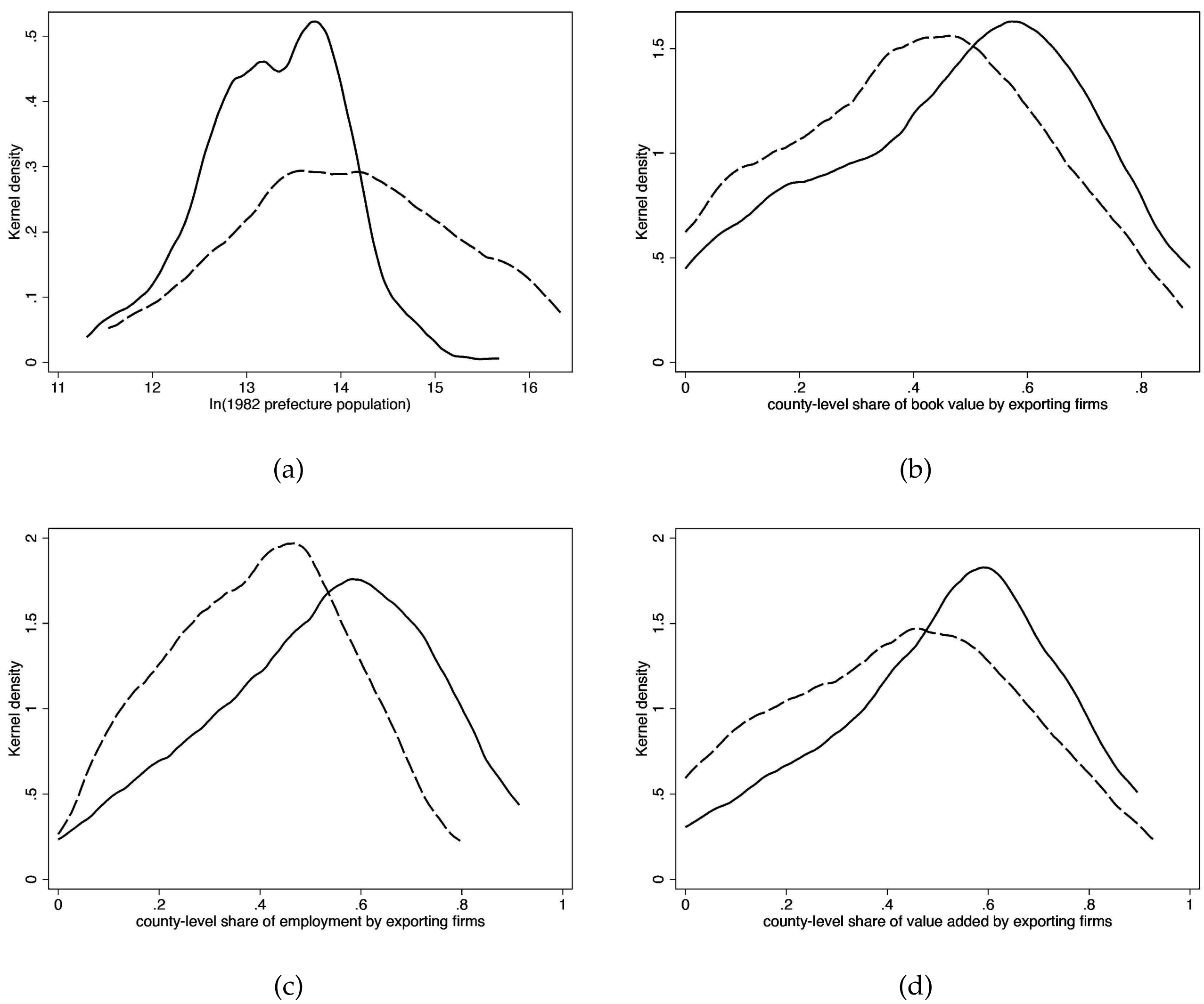
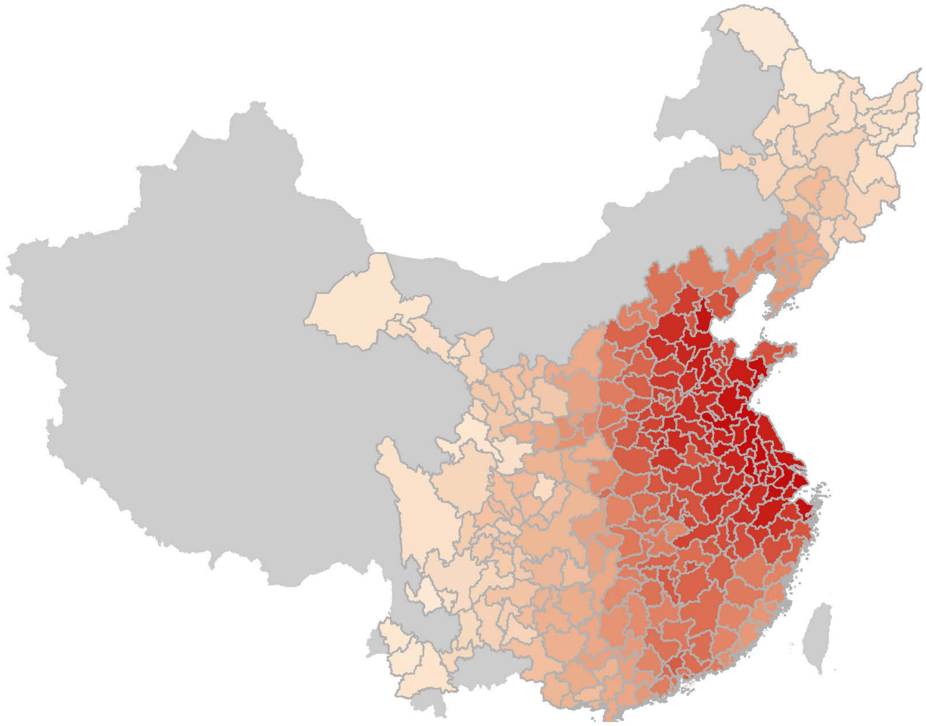
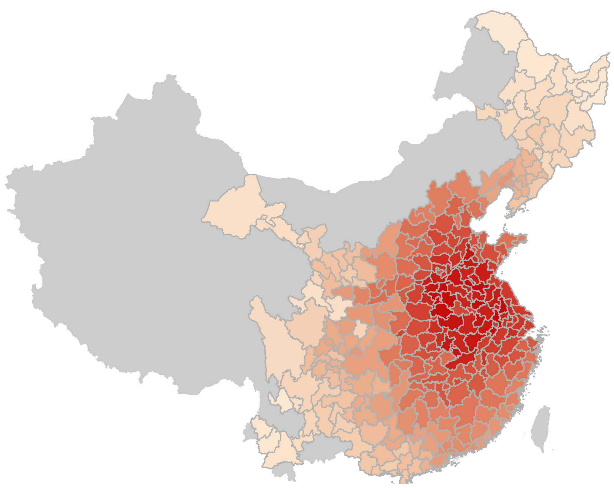


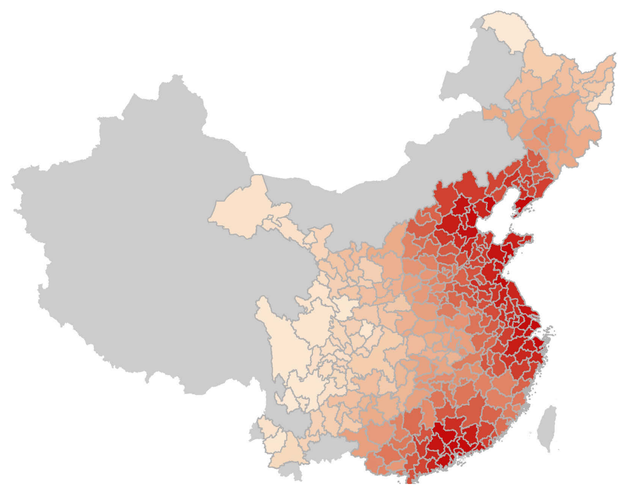
Figure 2: In all four panels, solid(dashed) line describes density for non-primate(primate) prefectures. Panel (a) shows the frequency of primate and non-primate prefectures by ln(1982 population). Panel (b)shows the frequency of NDZ counties in primate and non-primate prefectures on the y axis and the county-level share of book value by exporting firms. Panel (c) is like (b) but the x axis is county-level share of employment by exporting firms. Panel (d) is like (b) and (c) but the x axis is the county-level share of value added by exporting firms.



(a)

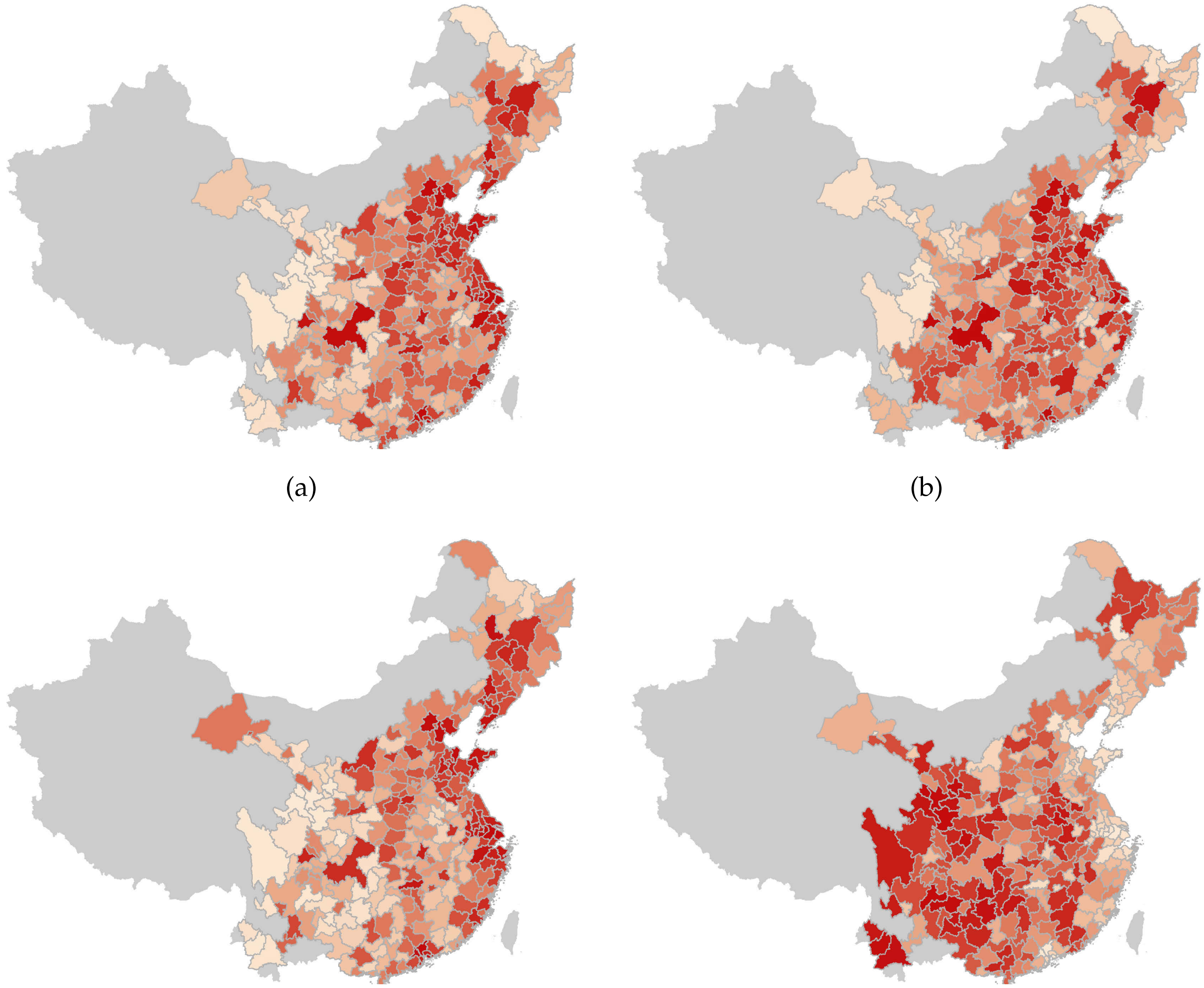


(b)



(c)

Figure 3: 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.



(a) (b)

(c) (d)

Figure 4: 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.

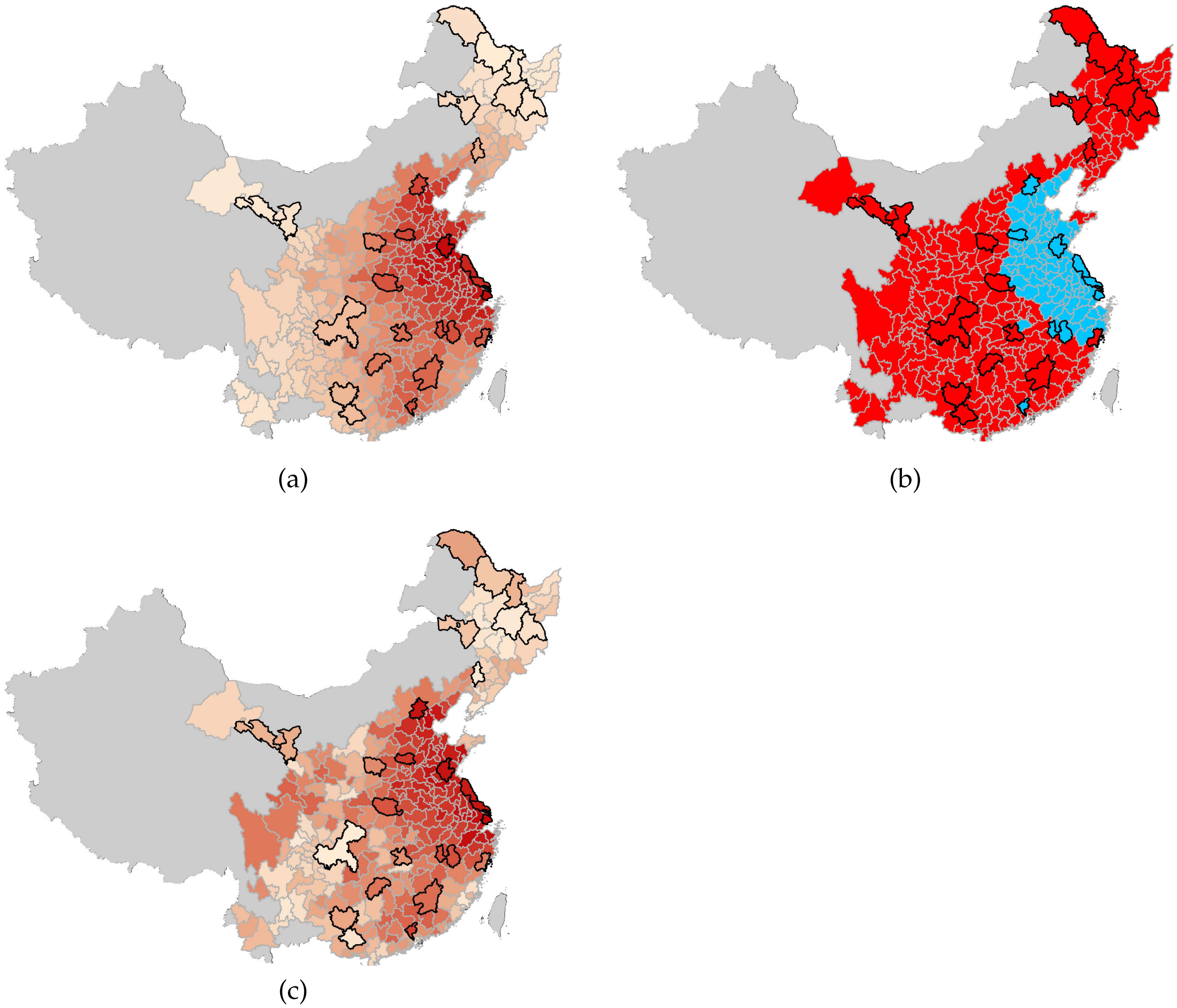


Figure 5: Counterfactual changes in GDP. Top is logs and bottom is levels. In the left column, colors indicate a prefectures ranking, darker colors indicate a larger increase in GDP under the counterfactual transportation network. In the right column, red indicates losers and blue indicates gainers. In all panels, highlighted prefectures are 'rank 1' as defined in the text.

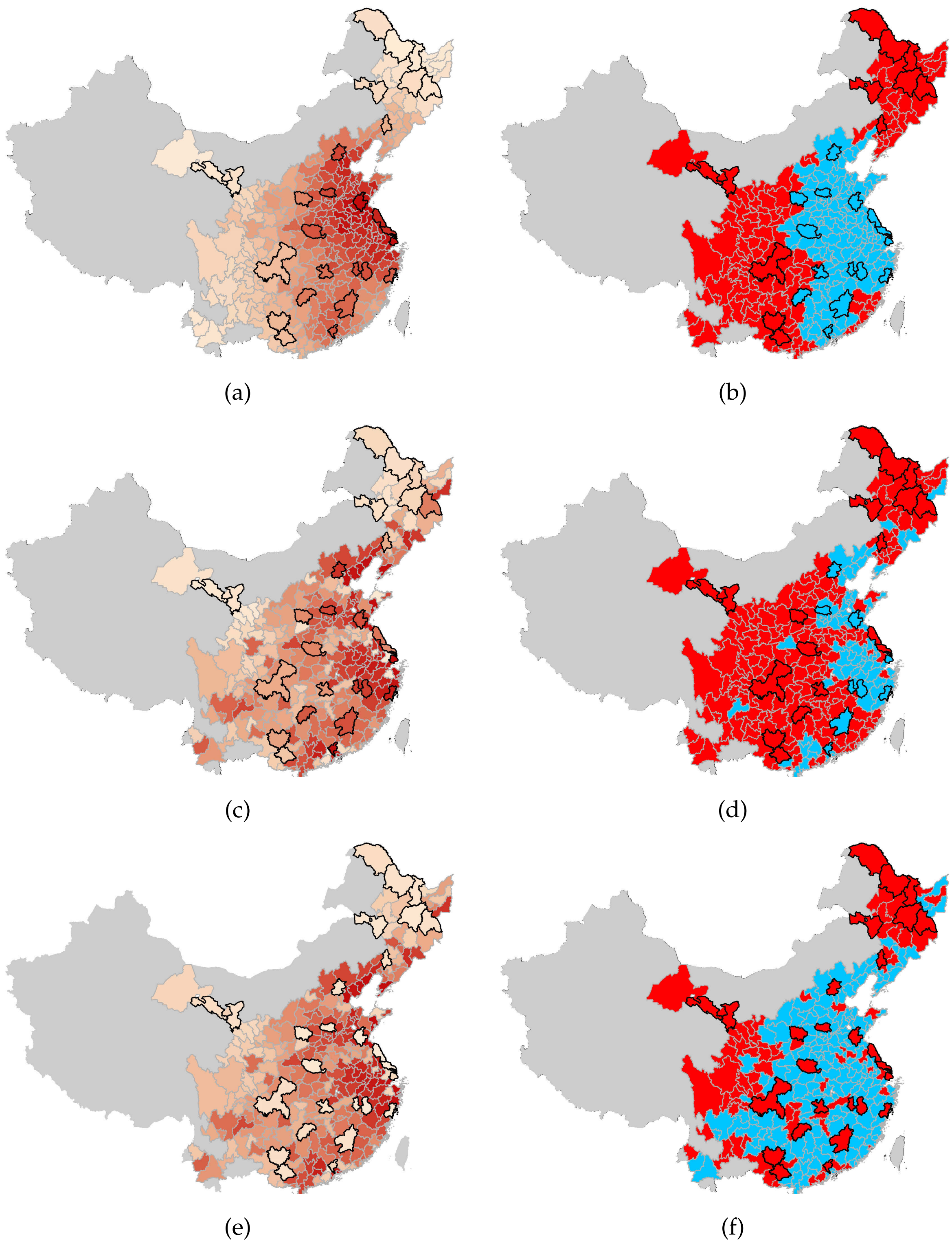


Figure 6: Counterfactual changes in logs of population. In the left column, colors indicate a prefectures ranking, darker colors indicate a larger increase in population under the counterfactual transportation network. In the right column, red indicates losers and blue indicates gainers. In all panels, 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.

The International Growth Centre (IGC) aims to promote sustainable growth in developing countries by providing demand-led policy advice based on frontier research.

Find out more about our work on our website www.theigc.org

For media or communications enquiries, please contact mail@theigc.org

Subscribe to our newsletter and topic updates www.theigc.org/newsletter

Follow us on Twitter @the_igc

Contact us
International Growth Centre,
London School of Economic and Political Science,
Houghton Street,
London WC2A 2AE

International Growth Centre

DIRECTED BY
LSE
UNIVERSITY OF OXFORD

FUNDED BY
UKaid from the British people

Designed by soapbox.co.uk