

The Distributional Impacts of Transportation Networks in China

Lin Ma Yang Tang *

May 2022

Abstract

This paper evaluates the dynamic and distributional impacts of transportation networks in China. We argue that the quality of roads and railroads vary substantially over time and space, and therefore binary measures of connectivity in the literature are inadequate. Instead, we construct a new panel dataset on China's road and railroad networks that accounts for quality differences. We measure the quality of infrastructure using the design speed that varies by vintage, rate, and the underlying terrain at the pixel level. We then build a dynamic spatial general equilibrium model that allows for multiple modes and routes of transportation and forward-looking migration to study the impacts of transportation. We show that the expansion of transportation networks significantly increases aggregate output and reduces spatial income inequality. Moreover, the return to better freight networks is high in the short-run but wanes in the long term. On the other hand, the return to better passenger networks could be negative in the short run but grows substantially over time.

Keywords: regional trade; migration; welfare; economic geography

JEL Classification: F1; F4; R1; O4

*Singapore Management University (linma@smu.edu.sg) and Nanyang Technological University (tangyang@ntu.edu.sg), respectively. Yang Tang acknowledges financial support from Singapore MOE-Tier 1 research fund(M4011853). We are solely responsible for the remaining errors.

1 Introduction

Over the past decades, developing countries worldwide have invested extensively in transportation networks.¹ While it is well-understood that the improvements in transportation infrastructure result in economic gains on average, the distributional impacts are much less clear.² From a theoretical perspective, better connectivity could either increase or decrease spatial inequality. On the one hand, improvements in trade frictions allow remote locations to enjoy better market access and reduce spatial inequality. On the other hand, however, better connectivity also improves factor mobility that drains resources away from remote locations. Moreover, the response of individuals to the changing infrastructure could be dynamic. While the benefits of better market access, such as lower prices, can be realized immediately, the changes in migration decisions could take years if individuals are forward-looking. Therefore, the distributional impacts of transportation are a quantitative question that needs to be answered through the lens of a dynamic general equilibrium model.

Such quantitative analysis is currently lacking in the context of China, a country that experienced one of the most spectacular transformations in transportation infrastructure in the past 40 years. The lack of quantitative analysis is due to many reasons, the most obvious of which is the lack of panel data that measure the evolution of transportation networks consistently across time and space. This paper aims to close these gaps. We first compile a new dataset that tracks the development of transportation networks and based on these data, we provide some first answers to the dynamic and distributional impacts of transportation networks in China.

On the empirical front, we construct a new panel dataset that documents the development of road and railroad networks in China each year between 1994 and 2017. In this dataset, we go beyond the binary measures of connectivity often used in the literature and instead measure the *quality* of road and railroad at the pixel level. Doing so allows for a meaningful comparison of infrastructure across time and space. Existing datasets often geo-reference the

¹For example, [Gurara et al. \(2018\)](#) estimated that the developing countries, on average, spend around 6 percent of GDP on infrastructure each year. The World Bank estimates that in the next decade, developing countries need to spend around 3.3 percent of GDP on transportation networks to meet the future demands for mobility, according to [Rozenberg and Fay \(2019\)](#).

²The literature of transportation costs on economic performance is extensive; see [Redding and Turner \(2015\)](#) for a summary of the literature.

highways and then rely on binary indicators of whether a highway exists on a certain pixel to measure the connectivity of locations.³ The implicit assumption behind these exercises is that all the highways are built with the same quality. However, this is not true. We show that the official engineering standards for roads and railroad design have improved substantially over the years. As a result, vintage highways constructed in the earlier years only allow for a fraction of travel speed compared to the recent ones. Moreover, conditional on the vintage and rate of the road, quality also varies across space. Due to engineering difficulties, the roads and railroads in the mountainous regions could be designed with half the speed compared to those built in the plain regions, leading to substantial inconsistency in cross-sectional comparison. In the empirical literature, the inconsistency introduced by standard revisions is widely seen as a significant obstacle to utilizing the time-series variations in transportation networks in China.⁴ The inconsistency of quality across space has not been noted by the existing literature.

To consistently measure the quality of the roads and railroads across time, we document the revisions in design codes over several decades using the official publications from the Ministry of Transportation in China. Each revision of the design code stipulates the design speed of roads and railroads by the rate and terrain: higher-rated roads such as “Highways” in the road system and “National I” in the rail system are often designed to allow for faster travel as compared to lower-rated ones, and rugged terrains such as hills and mountains impose limits on the design speed of infrastructure that trespasses them. To utilize this information, we first identify the year of construction and the rate of the road and railroad segments each year by cross-referencing a wide array of sources, including the *Transportation Yearbooks*, provincial map collections, and the *Chronicles of Railroad Construction* (Ma, 1983). With the information on the year of construction and rate, we can determine the

³For example, see the highway connectivity measures in Banerjee et al. (2012), Faber (2014), and Baum-Snow et al. (2020). The maps on the Chinese road system hosted by the China Data Center at the University of Michigan, the United States Geological Survey (USGS), ACASIAN Data Center at Griffith University, OpenStreetMap Project, and the State Bureau of Surveying and Mapping of China are all binary maps without any measure of quality.

⁴For example, Baum-Snow et al. (2020) observe that “...however, the growth and improvement of China’s road network was so dramatic that roads that were important enough to merit inclusion on the 1990 map probably bear little resemblance to roads that meet this standard in 2010, even if both roads receive the same designation in the legend. Thus, we are reluctant to exploit the time-series variation in our measures of highways.”

applicable design code for each segment. We then geo-reference the segments and determine the type of terrain at the pixel level using the terrain map of China. Combining all this information allows us to determine the “design speed”, which we interpret as the quality, at the pixel level for every road and railroad in China. From the travel speed, we then measure the distance between any two points by travel time, thus finishing our dataset. To our best knowledge, the resulting dataset is the first one that consistently measures the quality of transportation infrastructure in a country over time and across space, and certainly the first for China.

Besides the contribution in quality measures, our dataset also improves over the existing data in the following two aspects. First, differences in projection methods also lead to inconsistent geo-references in panel maps. As map publishers differ in projections methods, the same road segment could be inconsistently geo-referenced across the years if the source maps come from different publishers. To address this issue, we adopt an iterative procedure to construct the dataset to ensure that every pixel of the same road always has the same coordinates throughout the years. Secondly, we explicitly distinguish between freight and passenger travel on the railroad networks. This is particularly important in the context of China as the newly-constructed High-Speed Rail (HSR) system only allows for passenger traffic. The distinction is also important in welfare analysis, as the freight networks should only directly affect internal trade but not factor mobility, and the passenger networks the other way around.

Our data show that the expansions in transportation infrastructure between 1994 and 2017 have substantially reduced the average freight and passenger travel time between prefectures by 37 and 59 percent, respectively. Moreover, improvements in connectivity favor initially remote locations. For example, on the railroad network for freight transportation, while the well-connected prefectures in the initial year receive a modest improvement at around 20 percent, the remote ones could enjoy a cost reduction as high as 60 percent. The uneven improvement in connectivity is due to the asymmetry in the initial network. By 1994, the coastal cities were already well-connected, while the infrastructure in the inland regions was sparse. Under this circumstance, linking a remote prefecture to the extensive network along the coast reduces the travel time of that prefecture to all the coastal ones

in the existing network, therefore substantially reducing its average transport time. On the other direction, the additional access to the remote prefecture barely affects the average travel time of the coastal ones. However, one should not simply equate the remarkable gain in connectivity in the remote cities to the reduction in spatial inequality in economic variables for the reasons discussed above. While the improved connection with the large cities facilitates the goods flow, it also makes it easier for production factors to migrate away. The effect of the transportation improvements needs to be seen through the lens of a dynamic general equilibrium model.

To evaluate the general equilibrium impacts of transportation networks, we construct a dynamic spatial general equilibrium model in which both trade and migration flows are subject to bilateral frictions that depend on the observed freight and passenger infrastructure. We explicitly model route choices in the transportation network following [Allen and Arkolakis \(2022\)](#), and extend their framework to allow for multiple modes of transportation (road, railroad, waterway). To capture the mode choices, we introduce new parameters that govern the mode-specific usage costs and the time elasticities, and show that these parameters can be structurally estimated by matching the data moments on mode-specific usage. As we study the welfare impacts over a long period of time, we adopt a dynamic discrete choice framework as in [Caliendo et al. \(2019\)](#), and allow the migration decisions to be forward-looking. In the model, the location choice of individuals responds not only to the current state of passenger travel networks, but also to the anticipated changes in the future. We show that the forward looking behavior leads to rich dynamic responses of the economy to both the expansions of freight and passenger networks. On top of this framework we allow for the usual elements in the quantitative spatial literature, such as location-specific productivity and amenity, and agglomeration and congestion externality; we also model China-specific elements, such as the policy barriers to migration in the form of the hukou system.

We quantify the model to 291 prefectures in China with a mixture of methods. To structurally estimate the parameters on mode-specific costs and elasticities to travel time, we use the Generalized Method of Moments to match the data moments on traffic shares by mode in each prefecture between 1995 and 2017. Implementing this intuitive estimation procedure is not trivial. In particular, to generate the model-implied moments, we need to

solve the transition path in *levels* from 1995 towards a long-run steady-state beyond 2017. As a result, we cannot use the dynamic hat algebra introduced in [Caliendo et al. \(2019\)](#) and instead need to solve the model in levels. To do so, we invert the model to recover the location-specific productivity and amenities from the initial states in the year 1995 using the methods introduced in [Kleinman et al. \(2021\)](#). Together with the documented changes in transportation networks and migration policies, we solve all the endogenous variables in levels on the transition path and discipline the parameters of the model using the model-generated moments.⁵ Comparing to [Caliendo et al. \(2019\)](#), our quantification strategy is computationally more intensive. However, it offers two distinct advantages. First, it allows the researcher greater flexibility in matching the moments that are dependent on the *levels* of endogenous variables in quantification. Second, our quantification strategy has a lower data requirement as we do not need to observe the inter-prefecture trade flow. The second advantage is particularly helpful when working with countries such as China, in which the data on internal trade do not exist.

We evaluate the impacts of transportation networks by comparing two sets of simulations. In both simulations, we start with the initial conditions in 1995 and solve the transition path 50 years forward. In the “baseline” simulation, we use both the trade and migration costs matrices from the actual transportation networks in each year between 1995 and 2017, and fix the networks to their 2017 levels in all the subsequent years. In the “no-change” counterfactual, we fix both networks to the initial year in 1995. In other words, the transition path in the “baseline” economy includes all the actual changes and converges towards a steady-state as defined by the networks in 2017, while the “no-change” transition path converges towards a steady-state implied by the networks in 1995. Comparing the two sets of results reveals the impacts of the expansion of transportation networks. We highlight three findings here.

The first finding is that the growth of transportation networks between 1995 and 2017 improved aggregate output by around 60 percent in the long run. The aggregate gain comes from several sources: reducing trade frictions leads to the gains from trade, and

⁵Similar to [Caliendo et al. \(2019\)](#), we do not need to assume that the initial year, 1995, is in steady-state. Instead, we only assume that the economy was on a transition path towards some initial steady-state as defined by the transportation networks in that year.

more accessible passenger travel facilitates migration towards more productive locations. In the long run, the contributions of the trade and the migration channels are roughly equal.

Our second finding is that the dynamics of trade and migration channels are different. In particular, the return to trade costs reduction is high in the short-run but wanes in the long run; the return to migration costs reduction, on the other hand, could be negative in the short run but grows substantially over time. For example, by 2017, almost all the return to transportation networks come from better freight transportation, and only a negligible fraction comes from improved passenger travel. However, by 2044 at the end of the transition path, the contribution of freight and passenger networks are roughly equal. The return to freight transportation is immediate because firms and individuals respond to the changes in the trade network without any forward-looking concerns. However, in the long run, as trade cost reductions benefit remote locations more, they tend to keep the population in these remote locations that often lack productivity. In the long run, the equalizing effects of trade are a double-edged sword that leads to a slower gain in aggregate output.

The delayed return to migration liberalization is because the migration decisions are forward-looking. If individuals expect a better passenger travel network and lower policy restrictions in the future, they might delay their intended move despite the improvements in the current period. As a result, the population flow to the most productive prefectures is delayed as individuals prolong their stay in the less productive ones. The delay in migration leads to lower economic output in the short run. However, the return to migration liberalization is substantial and long-lasting in the long run, as it facilitates population movements toward more productive locations.

Lastly, our third finding is that the expansion of transportation networks significantly reduced spatial inequality. We measure spatial inequality using β -convergence of real wage, output, and population across prefectures. Without infrastructure improvements, spatial inequality in all three variables would have stayed the same without any regional convergence. In stark contrast, all three variables exhibit strong regional convergence in the baseline simulation that incorporates the actual changes in transportation networks. In other words, almost all the convergence in economic activities in China comes from the changes in transportation networks. We further decompose the distributional impacts and find that around

80 percent of the regional convergence in real wage and economic output is driven by the improvements in freight transportation. The migration channel played a minor role in reducing spatial inequality. Through the lens of the model, these results are expected as inter-region trade tends to benefit the remote locations while migration tends to draw workers away.

Our paper contributes to a large literature on the economic impacts of infrastructure improvements ([Baum-Snow, 2007](#); [Banerjee et al., 2012](#); [Faber, 2014](#); [Redding and Turner, 2015](#); [Donaldson and Hornbeck, 2016](#); [Qin, 2016](#); [Lin, 2017](#); [Xu, 2017](#); [Donaldson, 2018](#); [Baum-Snow et al., 2020](#); [Fan et al., 2021](#); [Alder et al., 2021](#)). In the context of China, our main contribution is two-fold. First, while the literature relies on binary measures of connection, we show that the quality of infrastructure differs over time and space and provide the first dataset to measure the quality of roads and railroads over an extended period. Second, we are the first to evaluate the impacts of transportation in a dynamic general equilibrium setup. Many reduced-form research document a “tunnel effect”, in which smaller and peripheral cities tend to lose from road expansions due to factor mobility ([Faber, 2014](#); [Qin, 2016](#); [Baum-Snow et al., 2020](#)). We complement this line of work by showing how spatial disparity depends on both goods and factor mobility. We further argue that the “tunnel effect” could be dominated by the positive impacts from the goods market due to the changes in market access.

Our work also contributes to the literature of quantitative spatial models ([Allen and Arkolakis, 2014](#); [Ahlfeldt et al., 2015](#); [Redding, 2016](#); [Tombe and Zhu, 2019](#); [Caliendo et al., 2019](#); [Kleinman et al., 2021](#); [Allen and Arkolakis, 2022](#)). The closest to our model are [Allen and Arkolakis \(2022\)](#) and [Caliendo et al. \(2019\)](#). Relative to [Allen and Arkolakis \(2022\)](#), we allow for multiple modes of travel and dynamic response of migration. We show that the welfare impacts of transportation networks could exhibit rich dynamics in short- and the long-run. In particular, the short-run return to infrastructure investment could be negative if individuals are forward-looking and expect future improvements. Relative to [Caliendo et al. \(2019\)](#), we introduce route choices in the dynamic migration framework. Despite the significant changes in the choice sets, we show that the migration decision still adopts a tractable solution. Moreover, we allow for time-varying migration costs and highlight that the expected changes in migration costs could lead to short-run losses from migration

liberalization.

The rest of the paper is organized as follows. Section 2 introduces the panel dataset on the transportation networks of China; Section 3 presents the model and Section 4 the quantification. Section 5 discusses the results and Section 6 concludes.

2 The Transportation Networks of China

In this section, we outline the construction of the panel dataset on transportation networks in China from 1994 to 2017 and refer the readers to Appendix B for details. We first briefly discuss the issues related to the digitization of the transportation atlas, then we discuss the underlying inconsistency from measuring the quality of road and railroad across time and space. Lastly, we describe the basic patterns on the evolution of the transportation networks over this period in China.

2.1 Measuring Binary Connectivity

The starting point of compiling the dataset is to collect the published transportation atlas for each year dated back as early as possible. We source the physical maps through several channels, such as libraries, used book dealers, and map collectors. We only choose the national-level atlas with a scale greater than 1:6 million, as smaller maps do not provide enough resolution for color identification. Out of the 24 years, we have obtained maps for ten years. To fill in the gap years, we resort to the annual *Collection of Provincial Transportation Maps* from various publishers, as well as the transportation yearbooks and chronicles for references. Table A.1 lists the information of the physical maps and other descriptive references in our collection.

One immediate challenge is that the same road is often projected differently across maps, leading to inconsistencies in geo-referencing. The variations in projection come from several sources. First and foremost, the publishers differ in projection methods. For example, while some publisher such as *Sino Maps* uses Albers projection with the reference point at the geographical center of China, other publishers such as *Guangzhou Publishers* uses the Lambert projection method with a reference point centered in the province of Guangdong.

Moreover, measurement errors also arise due to the noise in the designing, printing, and scanning the maps. As we aim to document the evolution of the transportation networks, we must ensure that the existing infrastructure is represented consistently over time.

We only geo-reference infrastructures based on their first appearance in the national maps to ensure consistency over time. In this way, the same road always has the same coordinates throughout the panel data. Of course, as the national maps only exist in sporadic years throughout the sample period, the year of the first appearance on the national map is not necessarily the year of the construction. To identify the year of the construction, we refer to the annually provincial map collections and other descriptive sources. More precisely, we first denote the 10 years in which the national maps are available as the “nodal years” and index them as $t_i, i = 1, 2, \dots, 10$. For example, in our data, the first nodal year is $t_1 = 1994$, and the second nodal year is $t_2 = 1996$. In each national map, we then extract four modes of transportation by color identification: first-rate road, highway, railway, and water. Table A.2 in the Appendix provides the details on the correspondence between map legends and the mode of transportation. The outcome is four binary maps by the modes of transportation in each of the ten nodal years.

By comparing the scanned maps between two consecutive nodal years t_i and t_{i+1} in a mode m , we can construct the annual maps for all the years in between. To do so, we treat each binary map as a connected graph and break the graph into “segments”, which are defined as a set of pixels between the branch points and endpoints of the graph (see Figure B.1 in the Appendix for an illustration). We then compare each segment in year t_{i+1} to that in year t_i to determine if it existed in t_i or it is newly constructed between t_i and t_{i+1} . Lastly, for each new construction, we determine the year of the construction by referencing the provincial maps, statistical yearbooks of transportation, railroad, as well as published chronicles on transportation construction in China.⁶ We repeat the above process for all the nodal years and all the modes.

At the end of this stage, our dataset identifies the year of the construction of each geo-

⁶Across the various reference sources, we first consult the yearbooks and the chronicles because they are official publications with the exact date of the construction. If a segment is not recorded in any yearbooks or chronicles, we then rely on visual identification from the provincial map collections to determine the year of construction.

referenced segment, from which one can trace the evolution of the connectivity by modes of transportation. Figure A.1 in the Appendix maps the evolution of the connectivity networks in several years. In empirical works, such as in Faber (2014) and Baum-Snow et al. (2020), connectivity is often measured as a binary indicator of whether roads exist on a pixel. Our dataset at this stage is already able to support such analysis. In the next part, however, we show that a binary measure of connectivity is insufficient, as the quality of the infrastructure varies significantly over time and space. To address this issue, we discuss the variation and measurement of the quality of the transportation networks in China.

2.2 Measuring Quality

The previous section created binary maps for all the transportation modes in all the years. However, it is hard to compare two roads or railroads across time and space due to the changes in the quality of the construction. Such variations are mostly driven by the evolution in the official engineering standards. Together with the rapid improvement in the capacity and capability of the civil engineering sector, the Chinese government has substantially lifted the standards of highway and railway construction over the years. As a result, a “highway” constructed in 1988 might only be appropriately classified as a “first-rate road” by the 2014 standards.

In addition to the temporal variation, quality also varies across space. Unlike the inconsistency over time, which are often recognized in the literature, many empirical works in this literature implicitly assume that the highways across the entire country are directly comparable in terms of traffic volume and speed. This is, unfortunately, not true. For example, under the 1988 construction standard, the highways in the eastern flood plains are designed for 120 km/h travel speed, and those in the mountainous regions are only built with a design speed of 60 km/h. One of the contributions in this paper is to carefully document the changes in the road and railway engineering standards over time and region and measure the quality of roads and railroads accordingly.

We use the “*design speed*” of roads and railroads as the measure of quality. The design speed is an ideal choice for two reasons. First, most parameters that govern the transportation engineering techniques are highly correlated with the design speed, and therefore this

single variable sufficiently reflects the quality of a road. For example, a road with a higher design speed is required to be built wider, straighter, and with stronger material. Secondly, from the design speed, one can directly compute the time cost of trespassing a pixel and, subsequently, the time cost of traveling between any two pixels on the map. The time cost of traveling can then be readily incorporated into a quantitative model of trade and migration, as discussed in detail in Section 3.

Lastly, to compute the design speed of each road and railroad pixel, we also need to measure terrains in China. The definition of the terrains comes from the *Land Regulations in Highway Engineering*, published by the Ministry of Transportation, as well as the United Nations definition of terrains from Kapos et al. (2000). In this section, we briefly discuss the measurement of road and railroad quality and refer the readers to Appendix B.2 for more details.

Roads Panel (a) of Table 1 summarizes the design speed of roads by the year and terrain of the construction. These stipulated design speeds are based on the *Technical Standard of Highway Engineering*, published by the Ministry of Transportation of China (MOT).⁷ Nine revisions have been made from 1951 to 2014, of which four are relevant for our sampling period: 1988, 1997, 2003, and 2014. Each revision defines the design speed for roads of different rates (highways, first, second, and third-rated roads) separately. We only focus on the highways and the first-rated roads identified in the previous step as described in Section 2.1. With the information on the year of the construction and the underlying terrain, we can determine the design speed for each road pixel.

In the earlier standards, the design speed of all roads depend on the terrain. For example, in the 1988 revision, the highways are constructed to allow for 120km/h design speed in the plains and low rolling hills (LRH), 100km/h in high hills, and 60 to 80km/h in the mountain regions. Similar rules apply to lower-rate roads with a lower design speed than highways at all terrains.⁸ Over time, the MOT has gradually loosened the dependency on terrain for

⁷Despite the name of “highway” in the title, the standards regulate all inter-city roads, including the highways (Gao Su Gong Lu) and normal roads (Yi Ban Gong Lu) from the first-rate to the fourth-rate. “Highway” in the title is the official translation of “Gong Lu”, which means inter-city roads. The “Urban roads” (Cheng Shi Dao Lu), the roads for intra-city transportation, are not regulated by these standards.

⁸For example, in 1988, the first-rate roads were designed for 100km/h in the plains and 60km/h in the

highway design but maintained the terrain-dependency for lower-rate roads. The changes in highway design are primarily due to two factors: 1) the rugged terrain is no longer the limiting factor of highway construction as China advances its engineering technology, and 2) the MOT realized that a highway with low design speed and capacity can hardly be upgraded to accommodate the rapidly increasing traffic volume. As a result, when 2003 revision was drafted, all the newly constructed highways were required to have a design speed of 120 km/h with very few exceptions. For the first-rate roads, which is more commonly addressed as “national roads”, the dependency on terrain is mainly maintained. In the plains and LRH, the design speed of the first-rate roads was always 100km/h throughout all four revisions, and in the hills and the mountains, the design speed was upgraded from 60km/h to 80km/h in the 2003 revision.

Railroads Panels (b) and (c) of Table 1 summarize the design speeds of mixed-use and freight-only railroads by year and construction terrain.⁹ Similar to the standards in highway engineering, the railroad designs depend on the intended usage, the rate of a railroad (National I-IV, Industrial I-III), and the underlying terrain. They have evolved significantly over time. In this section, we highlight several issues specific to the measurement of the railroad quality.

First, different from the road transportation, where passengers and goods share the same road, railroad designs vary by the intended usage and fall into three categories: passenger-only, freight-only, and mixed-use. The passenger-only railroads, including the High-Speed-Rail (HSR) system introduced in the last two decades, emphasize travel speed but cannot handle heavy loads. On the other hand, the freight-only railroads prioritize load capacity over the speed. Most of the railroads fall into the last category, the mixed-use railroads. This type of railroad allows for both passenger and freight traffic and often strikes a balance between speed and load. In addition, different from the road system where the legends in the maps identify the rate of the roads (see Table A.2), railroads of all usage and rates are often represented using a single legend in most of the maps. To identify the intended usage and rate

mountains. For the second-rate roads, the design speeds were lowered to 80 and 40km/h, respectively.

⁹These tables are based on various publications from the Ministry of Transportation and the Ministry of Railroads, such as the *Code for Design of Railway Line*, *Code for Design of Standard Railway Line for Industrial Firms*, and the *Code for Design of III and IV-rated Railway Line*.

Table 1: Design Speed (km/h) of Roads and Railroads by Time and Terrain

(a) Road Standards									
Revision	Plains	LRH	Hills	Mountains	Plains	LRH	Hills	Mountains	
Highways					First-Rate Roads				
1988	120	120	100	60	100	100	60	60	
1997	120	120	120	60	100	100	60	60	
2003	120	120	120	80	100	100	80	80	
2014	120	120	120	80	100	100	80	80	

(b) Railroad Standards, Mixed-Use									
Revision	Plains	LRH	Hills	Mountains	Plains	LRH	Hills	Mountains	
National I					National II				
1985	120	100	80	80	100	80	80	80	
1999	140	120	80	80	100	80	80	80	
2006	160	140	120	120	120	100	80	80	
National III					National IV				
1985	80	80	80	80	-	-	-	-	
1999	80	80	80	80	-	-	-	-	
2012	120	100	80	80	100	80	60	40	

(c) Railroad Standards, Freight-Only									
Revision	Plains	LRH	Hills	Mountains	Plains	LRH	Hills	Mountains	
Industrial I					Industrial II				
1987	70	70	70	70	55	55	55	55	
Industrial III									
1987	40	40	40	40					

Notes: this table summarizes the design speed of the highways and the first-rate roads by revisions of the *Technical Standard of Highway Engineering*, and that of the railroads by various revisions of railroad engineering. The “first-rate roads” is one tier below the highways, and is often considered interchangeable with the more commonly known term, “national roads” (Guo Dao). The definition of the terrains are provided in the *Land Regulations in Highway Engineering*, also published by the Ministry of Transportation. “LRH” refers to “Low Rolling Hills”. For more details, please refer to Appendix B.1.

of each segment, we again refer to the yearbooks and the chronicles of railroad construction. From these sources, we first recover the names of the railroads that each segment belongs to, we then identify the intended usage and the rate based upon their names.

Second, different from roads where the design speeds are fixed upon construction, the speed of existing railroads can also change over time due to continuous improvements in engine and route design. During our sampling period, the Ministry of Railroads implemented six waves of speed improvements (“Lu Wang Ti Su”) on the existing rail network, which increased the speed of old networks by as much as 30 percent in some cases.¹⁰ We incorporate such significant improvements in the dataset by referencing the yearbooks that provide the details of speed improvements of the major railroads.

Third, the following two groups of railroads are treated differently from the procedure above when we determine their design speeds. The first group is the HSR system, the passenger-only railroads. These railroads are designed with speeds much higher than the existing mixed-use networks and are not governed by the engineering standards behind Table 1. The second group is the long-haul freight-only railroads, such as the Dalian-Qinhuangdao, and Wazhai-Rizhao railroads, because Panel (c) of Table 1 only covers short-range industrial railroads. Long-haul freight trains typically travel at a faster speed than short-haul industrial railroads. We manually collect the design speed for each HSR and long-haul freight-only railroad from the yearbooks to address these issues.

Waterway To ensure consistency across the modes of transportation, we also measure the quality of waterway transportation by speed. As there are no significant improvements in sailing speed over the sampling period,¹¹ we assign a sailing speed of 22.8 km/h to all the waterway pixels in all the years. This sailing speed is the median sailing speed based on a sample of 5,112 vessels between the years 2012 and 2018 extracted from the Automatic Identification System (AIS) database.

¹⁰For example, the design speed of Longhai Railroad, built in the 1950s, was 120km/h before the first wave and increased to 160km/h after the sixth wave of speed improvements.

¹¹For example, the median sailing speed only varies between 20.7km/h and 23.2km/h across years in our sample.

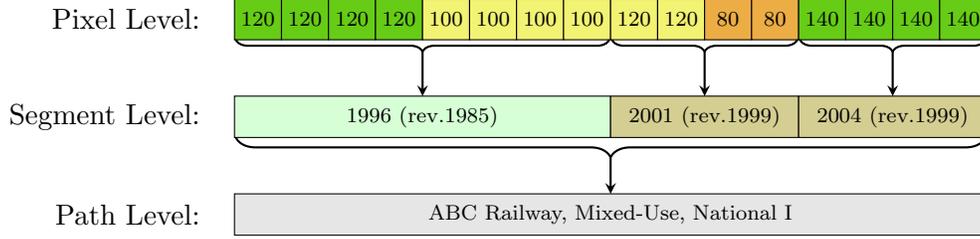


Figure 1: Structure of the Dataset

Notes: this diagram explains the structure of the dataset using a hypothetical railway, “ABC Railway”, as an example. At the path level, we identify the usage (mixed-use) and rate (National I), as well as the name of the path. This particular path contains three segments, built in 1996, 2001, and 2004, respectively. The segment built in 1996 was subject to the 1985 revision of *Code of Design of Railway Line*, while the two segments built in 2001 and 2004 were subject to the 1999 revision. Each segment contains a number of pixels that differ by terrain. ■ indicates “plains”, ■ indicates “low-rolling hills (LRH)”, and ■ indicates “hills”. The number in the pixel boxes is the design speed at each pixel in the unit of “km/h”. Within a segment, design speed differs due to terrains; within a terrain type, design speed differs due to the changes in applicable technical standards.

Structure of the Dataset Having implemented all the procedures described above, we reach a panel dataset that documents the evolution of the transportation networks. We briefly summarize the dataset’s three-layered structure in this part, conditional on a mode of transportation. Figure 1 illustrates the structure of the dataset using a hypothetical railway.

The outermost layer of the dataset is a “path”, which is a group of segments that form a known road or railroad. For example, a “path” could refer to a group of segments that form the Beijing-Shanghai Highway or the Longhai Railway. In addition to names, paths vary in two other dimensions: rate and usage. In the road dataset, a path could be rated as “Highway” or “First-Rated Roads”. Similarly, in the railroad dataset, a path can be classified as “National I, II,...”. The rate of the roads and railroads are then used to determine the applicable design codes later at the segment level. The railroad usage is also identified at the path level, which could be passenger-only, freight-only, or mixed-use. In contrast, all the road paths are mixed-use.

One layer below the “path” is the “segment”. Within a path, segments vary in the year of the construction. It is common for a path to be constructed over many years, and we can identify these based on variation at the segment level. The year of the construction then determines the revision of the technical standards to be applied when determining the design

speed of each pixel within the segment. As an example shown in Figure 1, the particular path contains three segments, built in 1996, 2001, and 2004, respectively. The segment built in 1996 was subject to the 1985 revision of *Code of Design of Railway Line*, while the two segments built in 2001 and 2004 were subject to the 1999 revision.

At the lowest level of the dataset are the “pixels” that form each segment. Within a segment, pixels differ in their underlying terrain. Together with the rate, usage, and year of the construction from the upper layers of the dataset, we can then determine the design speed of each pixel. In Figure 1, the color of the pixel indicates the terrain, and the number in each pixel states the design speed in the unit of “km/h”. Note that within a segment, design speed differs due to terrains; within a terrain type, design speed differs due to the changes in applicable technical standards.

Lastly, the waterway transportation data is a single-layer, cross-sectional dataset that identifies the navigable pixels on the map. We assign a common sailing speed to all the navigable pixels.

2.3 The Evolution of Transportation Networks in China

Based on the design speed at the pixel-level identified in the previous part, we compute the travel time between two prefectures, i, j at time t using the fast-marching algorithm. We denote the time cost matrix in the unit of hours as $\mathbf{T}_t^{m\chi}$, where the (i, j) th element of the matrix is the travel time from j to i under transportation mode m for traffic type χ . The mode of transportation, m , takes three values: $m = 1$ indicates the road, which further includes both the first-rate roads and the highways; $m = 2$ is the railroad, and $m = 3$ is the waterway network. We compute the travel time separately for two types of traffic denoted by χ , in which $\chi = \text{f}$ denotes the freight traffic, and $\chi = \text{p}$ the passenger traffic. Both traffic types travel on the road and waterway networks. On the railroad network, freight traffic cannot utilize the passenger railroads, and the passenger traffic cannot use the freight railroads. The estimated $\mathbf{T}_t^{m\chi}$ matrix offers a succinct summary of the evolution of transportation networks in China. Figures 2 to 4 present the basic findings. Three messages emerge.

The first message is that the expansion of transportation networks significantly reduced

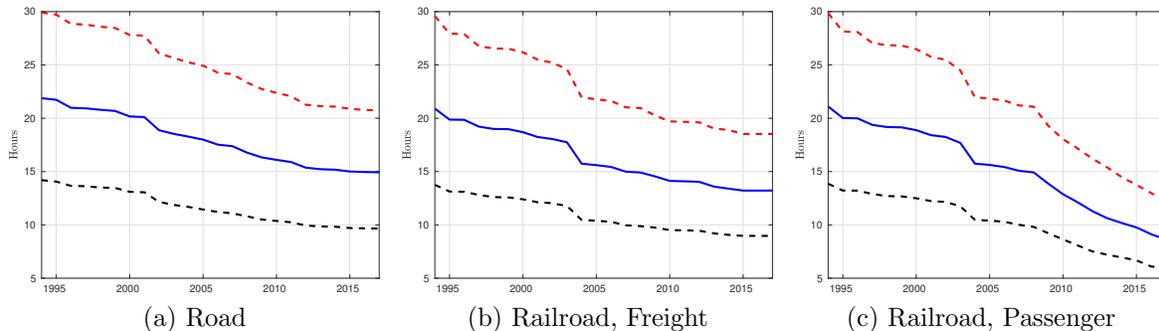


Figure 2: The Evolution of Travel Time

Notes: The figures present the average travel time in minutes across all prefecture-pairs by modes of transportation and types of traffic. In all the panels, the solid blue line in the middle is the median, the dashed red line is the 75th, and the dashed black line, the 25th percentile across all the prefecture-pairs. Note that the travel time on the road network is the same for freight and passenger traffic, and therefore we only present one set of figures under the sub-caption “road”. The figure for waterway transportation is omitted because it does not vary over time.

travel time in China. As shown in Figure 2, the median travel time between all prefecture-pairs on the road networks dropped by around 31 percent from close to 22 hours in 1994 to around 14 hours in 2017. The reduction in the travel time on the rail networks is more substantial, at 37 percent for freight travel and 59 percent for passenger travel. The significant reduction in the passenger travel time is mainly due to the expansion of the passenger-only HSR network. Moreover, the improvements in connectivity occurred across the board: at the 75th percentile of the travel-time distribution, we observe a reduction of similar magnitude, and at the 25th percentile, a slightly smaller reduction. The decreasing marginal returns in connectivity are intuitive. While a road connecting two remote prefectures reduces travel frictions substantially, the same might not be valid for a prefecture-pair closer to the center of the network.

The second message is that the reduction in the travel time varies substantially across prefecture-pairs, modes, and time. Figure 3 highlights this pattern by plotting the normalized travel time to three cities, Beijing, Guangzhou and Chengdu, from various origins. Some city pairs, such as Beijing-Shijiazhuang and Beijing-Tianjin in Panel (a), experienced a negligible reduction in travel time on the road network, and some other pairs, such as Beijing-Wulumuqi, saw a reduction of close to 30 percent. Along the temporal dimension, changes in travel costs also vary. For example, while the Beijing-Haerbin road connection

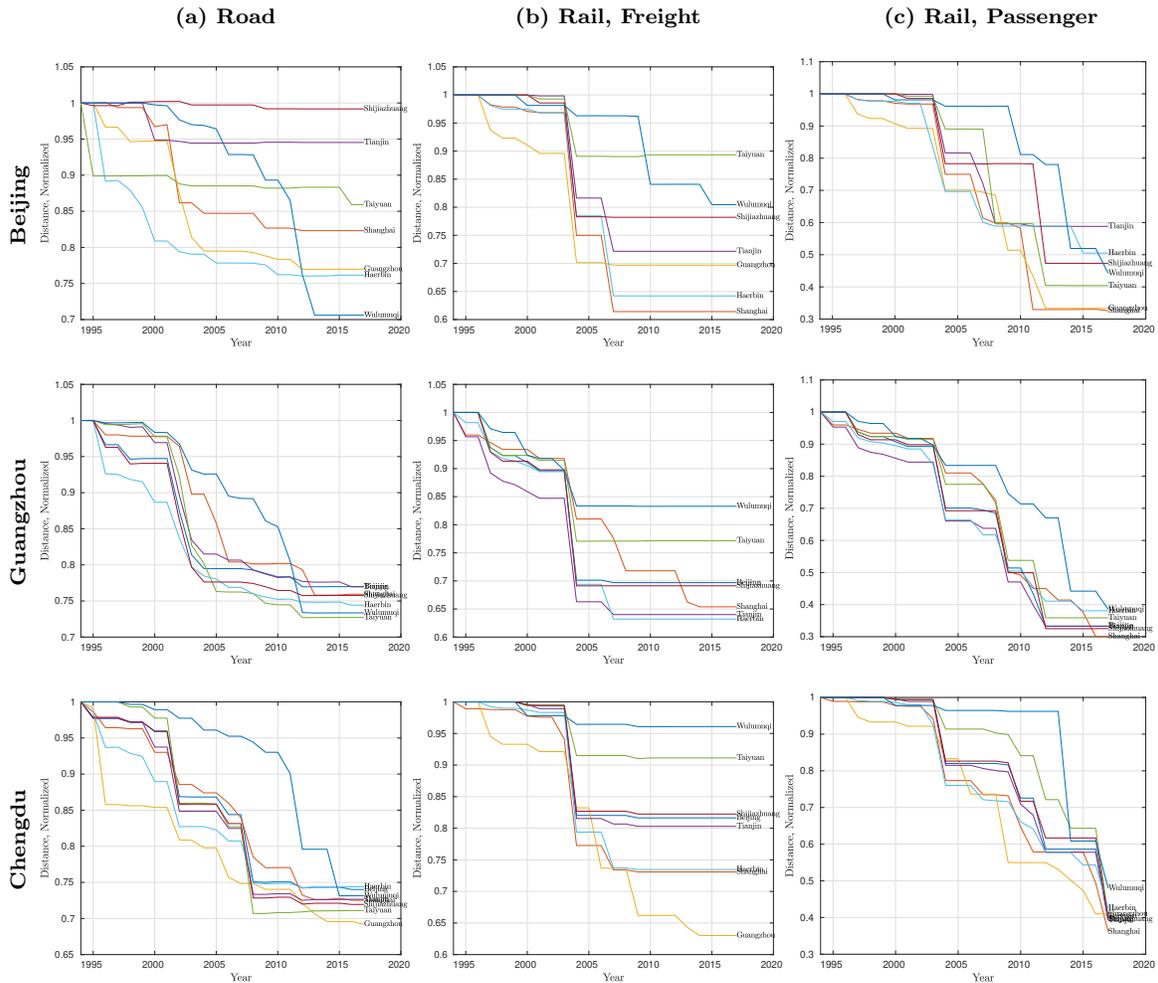


Figure 3: The Evolution of Transportation Networks, Distance to Selected Cities

Notes: The figures present the travel time between an origin city listed on the right axis of each figure and the selected cities on the column header, respectively, over time. The travel time is normalized to 1 in the initial year. The figure for waterway transportation is omitted because it does not vary over time.

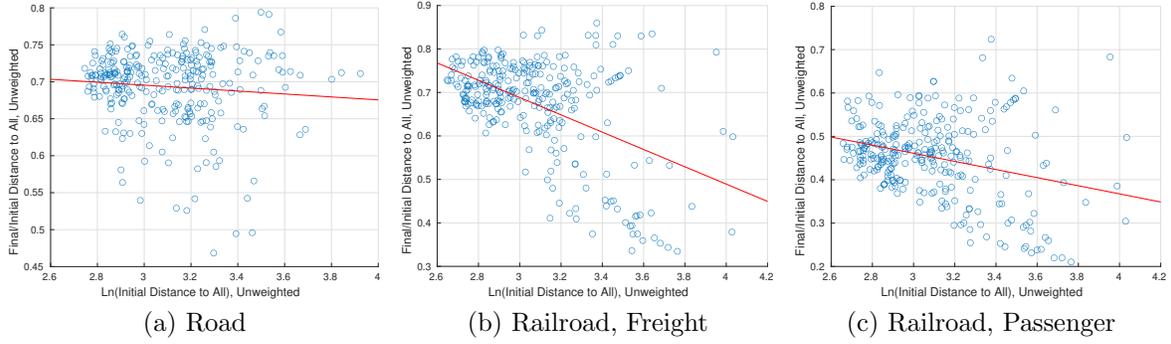


Figure 4: The Reduction in Transportation Costs v.s. Initial Position

Notes: this figure plots the reduction in travel time between 1994 and 2017 against the (logarithm of) the initial average travel time of each prefecture to all the other prefectures in 1994. Each dot represents a prefecture. The red line is the best linear fit. The figure for waterway transportation is omitted because it does not vary over time.

improved before 2000, those between Beijing and Wulumuqi only improved in 2012 after the construction of the Lanzhou-Wulumuqi Highway. The other panels of the same figure tell similar stories.

The last message is that those initially remote cities received a more considerable reduction in the travel time. This pattern is already reflected in the relatively milder reduction in travel time at the 25th percentile, as seen in Figure 2. Figure 4 plots the reduction in the average travel time from an origin prefecture between 1994 and 2017 against the initial average travel time in 1994. A negative relationship emerges in both the goods and passenger transportation in the rail networks, and a moderate negative relationship can also be seen in the road network. The uneven improvement in the connectivity is due to the asymmetry in the initial network layout, in which the coastal cities were already well-connected to each other while the infrastructure in the inland regions were sparse in 1995. Under this circumstance, linking a remote city to the extensive network along the coast will reduce the travel time of the inland city to all the coastal cities in the existing network, therefore substantially reducing its average transport friction. On the other direction, the additional access to the remote city barely affects the average connectivity of the coastal cities.

Before moving onto the structural model, we emphasize that one cannot simply equate the relative gain in connectivity in the remote cities to a relative gain in real wage and thus reducing spatial inequality. The improved connectivity might work as a double-edged sword.

While the improved connection with the large cities facilitates the goods flow, it also makes it easier for production factors to migrate away. The effect of the transportation improvements needs to be seen through the lens of a general equilibrium model, which we present in the next section.

3 The Model

The model is a combination of both economic geography model and urban model in AA(2022) with endogenous inter-city trade and migration pattern arising from endogenous trade and migration costs. The model also extends AA(2022) into a dynamic setting with multi-transportation modes.

3.1 Individuals and the Migration Decisions

The economy contains a mass $\bar{L} > 0$ of individual workers and $J > 1$ geographically segmented cities, indexed by $j = 1, 2, \dots, J$. An outside world (ROW), denoted as $j = 0$, trades with those J cities inside the country. Individuals can migrate between the $j = 1, \dots, J$ cities subject to frictions, but cannot move to or from ROW. Firms can trade between all the $J + 1$ locations subject to variable trade costs. Time, indexed by $t = 0, 1, \dots, \infty$ is discrete and infinite. Individuals living in city j at time t obtain flow utilities according to:

$$u_{jt} = \log(\phi_{jt} \cdot c_{jt}), \quad (1)$$

where c_{jt} is the consumption of a CES aggregation of intermediate goods indexed by ω defined over the real interval $[0, 1]$ with an elasticity of substitution denoted as η :

$$c_{jt} = \left(\int_0^1 (q_{jt}(\omega))^{\frac{\eta-1}{\eta}} d\omega \right)^{\frac{\eta}{\eta-1}}. \quad (2)$$

Individual also enjoys location-specific amenity that depends on an exogenously time invariant component $\bar{\phi}_j$ and the population size of the city at period t , L_{jt} :

$$\phi_{jt} = \bar{\phi}_j \cdot (L_{jt})^\beta, \quad (3)$$

where β captures the congestion elasticity. Individuals living in location j at period t receive the wage rate, w_{jt} , consume in city j with an ideal price index P_{jt} :

$$P_{jt} = \left(\int_0^1 (p_{jt}(\omega))^{1-\eta} d\omega \right)^{\frac{1}{1-\eta}}. \quad (4)$$

Migration At the end of the period, the individuals decide their locations in the next period, subject to migration frictions, which depend on both policy barriers and the transportation networks at time t . We interpret the policy element of the migration frictions as the hukou system in China that acts as entry barriers into location i , denoted as \bar{d}_{it} . The *direct* passenger transportation cost is a function of the passenger travel time that we have measured in Section 2. In particular, the direct cost of moving from j to i using mode $m = 1, 2, \dots, M$ at time t , denoted $d_{ijt}^{m\mathbb{P}}$, is:

$$d_{ijt}^{m\mathbb{P}} = a^{\mathbb{P}} \cdot a^{m\mathbb{P}} \cdot (T_{ijt}^{m\mathbb{P}})^{h^{\mathbb{P}}}, \quad (5)$$

where $a^{\mathbb{P}} > 0$ captures the average travel costs, $a^{m\mathbb{P}} > 0$ the mode-specific travel costs, and $h^{\mathbb{P}}$ is the elasticity of migration costs to travel time. We structurally estimate these parameters later. Denote the matrix of direct costs as $\mathbf{D}_t^{m\mathbb{P}}$ where the (i, j) th element is $d_{ijt}^{m\mathbb{P}}$. We follow [Allen and Arkolakis \(2022\)](#) to model the route choice of migrants. In order to move from j to i , the migrant decides a mode of transportation and then a *route* r^m of K steps. The route incurs a cumulative cost of $\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}}$, where $r_k^m (k = 1, 2, \dots, K)$ is the location of the k th-step in the route r^m .¹² Let \mathcal{R}_{ij}^m denote all the possible routes from j

¹²Different from the baseline setup in [Allen and Arkolakis \(2022\)](#) where the cumulative costs of a route takes the multiplicative format of $\prod_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}}$, we assume an additive format to be consistent with the Type-I Generalized Extreme Value shocks in the dynamic stochastic choice framework in [Artuc et al. \(2010\)](#) and [Caliendo et al. \(2018\)](#), which will be introduced later in this section. See Appendix D.1 in [Allen and Arkolakis \(2022\)](#) for more details of the additive travel costs.

to i . We abstract away from multi-modal routes so it is not allowed to switch transportation mode along a given route. In the context of inter-regional transportation in China, such an assumption is justified as multi-modal transportation only accounts for a negligible share of internal transportation.¹³

Given the policy barriers and the state of passenger transportation networks at time t , the individual residing in city j at time t observes a vector of idiosyncratic preference shocks toward each destination and route, $\{\varepsilon_{it,r^m}\}_{i=1,r^m \in \cup_{m=1}^M \mathcal{R}_{ij}^m}$, where ε_{it,r^m} is *i.i.d* across location, route, time, and individual. We assume that ε_{it,r^m} follows a Type-I extreme value distribution, with the following CDF:

$$F(\varepsilon) = \exp(-\exp(-\varepsilon - \bar{\gamma})),$$

where $\bar{\gamma}$ is the Euler's constant. Lastly, the individual discounts future with a rate δ . Taking into account all the components described above, the migration decision for an individual living in j at time t can be formally written as

$$v_{jt} = \log\left(\phi_{jt} \frac{w_{jt}}{P_{jt}}\right) + \max_{i,r^m \in \cup_{m=1}^M \mathcal{R}_{ij}^m} \left\{ \delta \mathbb{E}[v_{i,t+1}] - \bar{d}_{it} - \sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}} + \kappa \cdot \varepsilon_{it,r^m} \right\}, \quad (6)$$

where v_{jt} is the lifetime utility of the individual currently living in location j , and the expectation is taken over the future realizations of the idiosyncratic preference shock.

Comparing to the dynamic migration framework in [Caliendo et al. \(2019\)](#), we differ in two aspects. First, [Caliendo et al. \(2019\)](#) assumes time-invariant migration costs, and we allow the migration costs to be time-variant to capture the changes in transportation networks and policy barriers. Second, we embed the route-choice problem from [Allen and Arkolakis \(2022\)](#) into the dynamic migration model, so the individuals in our model are not choosing a destination prefecture, but a pair of destination and route at the same time. Despite the changes in the optimization problem, we show that the dynamic migration problem adopts a similar solution as compared to [Caliendo et al. \(2019\)](#) and thus the model can be solved using the standard iterative algorithms.

¹³For example, [Huang and Mu \(2018\)](#) reports that by 2015, only 2.9 percent of China's internal freight transportation is multi-modal.

Let $V_{jt} \equiv \mathbb{E}[v_{jt}]$ denote the expectation of the continuation value, we show in Appendix C.1 that V_{jt} takes the following form:

$$V_{jt} = \log \left(\phi_{jt} \frac{w_{jt}}{P_{jt}} \right) + \kappa \log \left(\sum_{i=1}^J \exp(\delta V_{i,t+1} - \lambda_{ijt})^{1/\kappa} \right), \quad (7)$$

where λ_{ijt} is the *expected* travel costs across all possible modes and routes from j , conditional on moving to i :

$$\lambda_{ijt} = \bar{d}_{it} - \kappa \log \left[\sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \exp \left(- \frac{\left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}} \right)}{\kappa} \right) \right]. \quad (8)$$

The expected life-time utility, V_{jt} , contains two parts: the flow utility in the first term, and the option value of living in location j at period t as summarized in the second term. As standard in the dynamic discrete choice literature, we can also derive the migration probability from j to i in period t as:

$$\mu_{ijt} = \frac{\exp(\delta V_{i,t+1} - \lambda_{ijt})^{1/\kappa}}{\sum_{i'=1}^J \exp(\delta V_{i',t+1} - \lambda_{i'jt})^{1/\kappa}}. \quad (9)$$

In the above expression, κ can also be interpreted as the migration barrier elasticity of bilateral migration. Lastly, the population distribution in the next period is:

$$L_{i,t+1} = \sum_{j=1}^J \mu_{ijt} L_{jt}. \quad (10)$$

Solution of λ_{ijt} Before moving onto the production and the trade part of the model, we first briefly discuss the solution of the expected travel cost, λ_{ijt} , and refer the readers to Appendix C.2 for more details. The definition in equation (8), while intuitive, is hard to compute as the cardinality of the set \mathcal{R}_{ij}^m is infinite and countable. We follow the steps in [Allen and Arkolakis \(2022\)](#) to enumerate all the possible routes, and thus compute λ_{ijt} with Leontief inversion.

In particular, we denote the adjacency matrix as $\mathbf{F}_t^{m\mathbb{P}}$, in which the (i, j) th element is $F_{ijt}^{m\mathbb{P}} = \exp \left(- \frac{d_{ijt}^{m\mathbb{P}}}{\kappa} \right)$. We further denote $(\mathbf{F}_t^{m\mathbb{P}})^K$ as the adjacency matrix raised to the matrix

power K . In Appendix C.2 we show that one can enumerate all the routes with length K in the format of matrix power and rewrite λ_{ijt} as:

$$\lambda_{ijt} = \bar{d}_{it} - \kappa \log \left\{ \sum_{m=1}^M \sum_{K=0}^{\infty} (\mathbf{F}_t^{m\mathbb{P}})_{ij}^K \right\},$$

where $(\mathbf{F}_t^{m\mathbb{P}})_{ij}^K$ is the (i, j) th element of the adjacency matrix raised to the power K . This step expresses λ_{ijt} as function of $\mathbf{F}_t^{m\mathbb{P}}$, which is observable conditional on κ . To proceed further, define $\mathbf{B}_t^{m\mathbb{P}}$ as the Leontief Inverse of $\mathbf{F}_t^{m\mathbb{P}}$:

$$\mathbf{B}_t^{m\mathbb{P}} \equiv (\mathbf{I} - \mathbf{F}_t^{m\mathbb{P}})^{-1} = \sum_{K=0}^{\infty} (\mathbf{F}_t^{m\mathbb{P}})^K.$$

We then express λ_{ijt} as a simple function of the policy barrier \bar{d}_{it} and the transportation network summarized in $\mathbf{B}_t^{m\mathbb{P}}$:

$$\lambda_{ijt} = \bar{d}_{it} - \kappa \log \left(\sum_{m=1}^M b_{ijt}^{m\mathbb{P}} \right), \quad (11)$$

where $b_{ijt}^{m\mathbb{P}}$ is the (i, j) th element of the matrix $\mathbf{B}_t^{m\mathbb{P}}$.

3.2 Production and Trade

Production The production side of the economy follows [Eaton and Kortum \(2002\)](#): the market structure is perfectly competitive and every city is able to produce every variety $\omega \in [0, 1]$. The production function for variety ω in city j at time t takes the form:

$$q_{jt}(\omega) = A_{jt} \cdot \ell_{jt},$$

where ℓ_{jt} is the labor input. A_{jt} is the city-specific productivity that depends on the fundamental productivity of the location, \bar{A}_j , as well as a time-varying part that is a function of the population size in the city to capture the agglomeration force with an elasticity of α :

$$A_{jt} = \bar{A}_j \cdot (L_{jt})^\alpha. \quad (12)$$

Internal Trade Costs We model the internal trade costs following [Allen and Arkolakis \(2022\)](#), similar to what we did to passenger travel costs. Conditional on a mode of transportation, m , the *ad valorem* cost of moving directly from j to i at time t is $d_{ijt}^{m\text{f}} \geq 1$. Further denote the matrix of $d_{ijt}^{m\text{f}}$ as $\mathbf{D}_t^{m\text{f}} = [d_{ijt}^{m\text{f}}]$. We assume that the direct cost of goods transportation is a function of the observed time cost of freight traffic discussed in Section 2, $\mathbf{T}_t^{m\text{f}}$:

$$d_{ijt}^{m\text{f}} = \exp \left(a^{\text{f}} \cdot a^{m\text{f}} \cdot (T_{ijt}^{m\text{f}})^{h^{\text{f}}} \right), \quad (13)$$

where $a^{\text{f}} > 0$ governs the average internal trade costs, $a^{m\text{f}} > 0$ the mode-specific cost, and h^{f} the elasticity of trade frictions to transportation time. Different from equation (5), the direct trade costs are ad valorem, and therefore require an exponential transformation to ensure $d_{ijt}^{m\text{f}} \geq 1$. Conditional on mode m , moving from i to j through a route r^m of K steps incurs a multiplicative cost of $\prod_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\text{f}}$, where r_k^m ($k=1, 2, \dots, K$) is the location of the k th-step in the route r^m . The cumulative costs here adopt a multiplicative format so that they are consistent with the tradition of Frechet shocks in the trade literature. Consumers in i face the following price if they source variety ω from location j through route r^m :

$$p_{ijt, r^m}(\omega) = \frac{w_{jt} \prod_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\text{f}}}{A_{jt} z_{ijt, r^m}(\omega)},$$

and individuals source from the cheapest location-route. We assume that $z_{ijt, r^m}(\omega)$ is i.i.d. Frechet shock across time, location, route, and variety and has a shape parameter θ . From these assumptions, it is straightforward to show that the probability that individuals in i source from j along route r^m is:

$$\pi_{ijt, r^m} = \frac{(w_{jt}/A_{jt})^{-\theta} \left(\prod_{k=1}^K \left(d_{r_{k-1}^m, r_k^m, t}^{m\text{f}} \right)^{-\theta} \right)}{\sum_{n=0}^J (w_{nt}/A_{nt})^{-\theta} \left[\sum_{m=1}^M \sum_{(r^m)' \in \mathcal{R}_{in}^m} \prod_{k=1}^K \left(d_{(r^m)'_{k-1}, (r^m)'_k, t}^{m\text{f}} \right)^{-\theta} \right]}, \quad (14)$$

and the expected transportation cost between i and j is:

$$\tau_{ijt} = \left[\sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \prod_{k=1}^K \left(d_{r_{k-1}^m, r_k^m, t}^{m\text{f}} \right)^{-\theta} \right]^{-\frac{1}{\theta}}. \quad (15)$$

By enumerating all the possible routes under mode m by length, we can re-write the above equation as:

$$\tau_{ijt}^{-\theta} = \sum_{m=1}^M \sum_{K=1}^{\infty} \left(\mathbf{F}_{ijt}^{m\text{f}} \right)^K,$$

where $\mathbf{F}_t^{m\text{f}} \equiv \left[\left(d_{ijt}^{m\text{f}} \right)^{-\theta} \right]$ is the adjacency matrix of mode m , $F_{ijt}^{m\text{f}}$ is the (i, j) th element of the matrix, and $\left(\mathbf{F}_t^{m\text{f}} \right)^K$ is the matrix raised to the power K . The power sequence can be evaluated as the Leontief Inverse:

$$\sum_{K=1}^{\infty} \left(\mathbf{F}_t^{m\text{f}} \right)^K = \left(\mathbf{I} - \mathbf{F}_t^{m\text{f}} \right)^{-1} \equiv \mathbf{B}_t^{m\text{f}}.$$

Denote the (i, j) th element of the matrix $\mathbf{B}_t^{m\text{f}}$ as $b_{ijt}^{m\text{f}}$, we can then write τ_{ijt} as:

$$\tau_{ijt} = \left(\sum_{m=1}^M b_{ijt}^{m\text{f}} \right)^{-\frac{1}{\theta}}.$$

Gravity Equation Aggregate equation (14) over m and \mathcal{R}_{ijt}^m leads to the probability that location i sources from j across all possible routes and modes of transportation:

$$\begin{aligned} \pi_{ijt} &= \sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ijt}^m} \pi_{ijt, r^m} \\ &= \sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ijt}^m} \frac{(w_{jt}/A_{jt})^{-\theta} \left(\prod_{k=1}^K \left(d_{r_{k-1}^m, r_k^m, t}^{m\text{f}} \right)^{-\theta} \right)}{\sum_{n=0}^J (w_{nt}/A_{nt})^{-\theta} \left[\sum_{m=1}^M \sum_{(r^m)' \in \mathcal{R}_{in}^m} \prod_{k=1}^K \left(d_{(r^m)'_{k-1}, (r^m)'_k, t}^{m\text{f}} \right)^{-\theta} \right]} \\ &= \frac{(w_{jt}/A_{jt})^{-\theta} \tau_{ijt}^{-\theta}}{\sum_{n=0}^J (w_{nt}/A_{nt})^{-\theta} \tau_{int}^{-\theta}}. \end{aligned} \quad (16)$$

The above also implies the expenditure share of city i on the goods from city j as:

$$X_{ijt} = X_{it}\pi_{ijt} = w_{it}L_{it} \frac{(w_{jt}\tau_{ijt})^{-\theta} (A_{jt})^\theta}{\sum_{n=0}^J (w_{nt}\tau_{int})^{-\theta} (A_{nt})^\theta}. \quad (17)$$

Summing the above equation over i leads to the total income in location j :

$$w_{jt}L_{jt} = \sum_{i=0}^J \frac{(w_{jt}\tau_{ijt})^{-\theta} (\bar{A}_j L_{jt}^\alpha)^\theta}{\sum_{n=0}^J (w_{nt}\tau_{int})^{-\theta} (A_n L_{nt}^\alpha)^\theta} w_{it}L_{it}. \quad (18)$$

Conditional on population $\{L_{jt}\}$, we can back out $\{w_{jt}\}$ from the above system of equations.

Lastly, the equilibrium price index in city j is:

$$P_{jt} = \Gamma\left(\frac{\theta + 1 - \eta}{\theta}\right) \left(\sum_{i=0}^J (w_{it}\tau_{jit})^{-\theta} (A_{it})^\theta\right)^{-1/\theta}, \quad (19)$$

where $\Gamma(\cdot)$ is the standard Gamma function.

3.3 The Equilibrium

Define the time-invariant fundamentals as $\bar{\Omega} = \{\bar{A}_j, \bar{\phi}_j\}$, the time-variant fundamentals as $\Omega_t = \{\bar{d}_{jt}, \mathbf{D}_t^{m\text{p}}, \mathbf{D}_t^{m\text{f}}\}$, and the sequence of endogenous variables as $\Upsilon_t = \{w_{jt}, L_{jt}, V_{jt}, \mu_{ijt}\}$.

We first define the sequential equilibrium, and then the steady state.

Sequential Competitive Equilibrium Conditional on $\{\bar{\Omega}, \Omega_t\}$ and the initial population $\{L_{j0}\}$, a sequential competitive equilibrium of the model is a sequence of Υ_t that solves the following equilibrium conditions:

1. Individuals maximize their life-time utility ((7)) by choosing a sequence of locations so that equations (9) and (10) hold.
2. Firms maximize their profits in each period and trade balances, so that equation (18) holds.

Steady state A steady-state of the economy is an equilibrium where both economic fundamentals and endogenous variables are time invariant.

4 Quantification

We quantify the model to 291 prefecture-level cities in China, plus one additional location representing the rest of the world (ROW). The 291 cities in our sample are the largest common sample available in which we can observe both the population and economic output dated back to 1995. We set the time at the yearly frequency, and $t = 1$ corresponds to the year 1995 in the data. In the rest of this section, we first expand the model to include the ROW and then briefly outline the estimation procedure of all the parameters in the model.

4.1 Rest of the World (ROW)

We model ROW as a single location indexed as $j = 0$ with population $L_{0t} = L_0, \forall t$. Recall that goods and people can move between the Chinese prefectures with friction, but migration between China and the ROW is not allowed. Under these assumptions, we need to specify the production and the trade decision made by ROW but not the individual migration decisions.

As the ROW does not correspond to any single geographical location, we ignore the agglomeration forces and assume the following production function:

$$q_{0t}(\omega) = \bar{A}_0 \cdot \ell_{0t}, \quad (20)$$

where \bar{A}_0 is the fundamental productivity of ROW to be calibrated later.

We model the trade costs between ROW and Chinese prefectures as follows. We first identify the coastal prefectures that directly import and export from the international markets using the transaction-level data from the Chinese Customs and designate these prefectures as the “port cities” (see Table A.3 in the Appendix for the list of these prefectures). We assume an identical and symmetric iceberg trade cost between any port city and the ROW and denote it as τ^* . The iceberg trade costs between a non-port prefecture i and the ROW are then assumed to be $\tau_{i,ROW,t} = \tau_{ROW,i,t} = \tau^* \cdot \tau_{i,j_t(i),t}$, in which $j_t(i)$ is the nearest port city to prefecture i in year t , as measured by τ_{ij_t} . The nearest port city could vary over time due to the changes in the transportation networks.

4.2 External Parameters

We choose the value for following parameters from the literature. The agglomeration elasticity, $\alpha = 0.1$ follows [Redding and Turner \(2015\)](#). Both the congestion elasticity $\beta = -0.3$ and the trade elasticity, $\theta = 6.2$ come from [Allen and Arkolakis \(2022\)](#). The migration elasticity, $\kappa = 2.02$, is taken from [Caliendo et al. \(2019\)](#) based on their estimates at the annual frequency. The elasticity of substitution in the utility function, η , does not affect any quantitative results after normalization, because it only scales the price index and thus the welfare level in the equilibrium. We set $\eta = 6$, a value in the middle of the commonly estimated range as reported in [Anderson and van Wincoop \(2004\)](#).

Policy Barrier, \bar{d}_{jt} We map the entry barrier in the migration frictions, \bar{d}_{jt} , to the hukou reform index constructed in [Fan \(2019\)](#). In that paper, the hukou reform index summarizes the gradual relaxation of the entry barriers at the prefecture-level from 1998 to 2010, based on official announcements. We follow its strategy and extend the index to the 291 prefectures in our sample between 1994 and 2017 using the same data source. Denote the observed hukou index in prefecture j at time t as $k_{jt} \geq 0$, we then model $\bar{d}_{jt} = \exp(\psi \cdot k_{jt})$ and estimate ψ in the next part.

4.3 Estimation and Joint Calibration

While the parameters listed in the previous section were pinned down outside of the model, we estimate and calibrate all the other parameters based on the model solutions on the transition path. We adopt a two-layer quantification strategy. On the outer layer, we use the Generalized Method of Moments (GMM) to estimate the parameters that govern the costs and time elasticity of the transportation networks, $\Theta^x = \{[a^{m^f}, a^{m^p}]_{m=1}^3, h^f, h^p\}$. In the inner-layer during the GMM estimation, conditional on a vector of Θ^x , we jointly calibrate all the other parameters: $\Theta = \{\bar{A}_j, \bar{\phi}_j, a^f, a^p, \tau^*, \psi\}$.

Table 2: Parameters

(a) External Calibrated

name	value	source	notes
α	0.1	Redding and Turner (2015)	agglomeration elasticity
β	-0.3	Allen and Arkolakis (2022)	congestion elasticity
θ	6.2	Allen and Arkolakis (2022)	trade elasticity
κ	2.02	Caliendo et al. (2019)	migration elasticity
η	6.0	Anderson and van Wincoop (2004)	elasticity of substitution

(b) Joint Calibrated

name	value	target	notes
$\{\bar{A}_j\}$	-	output in 1995	fundamental productivity
$\{\phi_j\}$	-	population in 1995	fundamental amenity
a^{f}	0.815	internal-trade-to-gdp ratio, 2002	average internal trade costs
a^{P}	5.823	average annual stay rate between 2000 and 2005	average migration costs
τ^*	1.614	average export-to-gdp ratio, 2000 to 2005	international trade barrier
ψ	0.039	average annual stay rate between 2010 and 2015	elasticity of \bar{d}_{it} to the hukou reform index

(c) Estimated

name	value	s.e.	notes
$a^{1,\text{f}}$	0.984	0.078	road costs, freight
$a^{2,\text{f}}$	1.151	0.085	rail costs, freight
h^{f}	0.111	0.034	time elasticity, freight
$a^{1,\text{P}}$	0.884	0.085	road costs, passenger
$a^{2,\text{P}}$	1.275	0.093	rail costs, passenger
h^{P}	0.320	0.051	time elasticity, passenger

Notes: this table reports the results of calibration and estimation. The parameters in Panel (a) are calibrated outside of the model. Those in Panel (b) are jointly calibrated on a transition path from the year 1995 to the long-run steady state, and those in Panel (c) are estimated using the GMM by simulating the same transition path. The standard errors are computed using Gauss-Newton Regression.

4.3.1 Inner Layer: Joint Calibration

We first describe joint calibration in the inner layer conditional on a vector of Θ^χ from the GMM. At this stage, with a guess of Θ , we have all the information to solve the transition path in levels. In particular, we solve the model for 50 years from the year 1995 which corresponds to $t = 1$. We first compute $\Omega_t = \{\bar{d}_{jt}, \mathbf{D}_t^{m\text{P}}, \mathbf{D}_t^{m\text{f}}\}$ from the observed data $\{k_{jt}, \mathbf{T}_t^{m\text{P}}, \mathbf{T}_t^{m\text{f}}\}$ and the parameters specified in $\{\Theta^\chi, \Theta\}$ for the years before 2017, at which point $t = 23$. We then assume that the time-varying fundamentals remain at the same levels as in the year 2017 for all the subsequent years, so that $\Omega_t = \Omega_{23}, \forall t > 23$. We do

not assume that the year 1995 is in an initial steady state. Instead, we assume that the initial year is somewhere along a transition path to a future steady state, and compute the transition path using a shooting algorithm. With the solution to the transition path, denoted as $\Upsilon_t = \{w_{jt}, L_{jt}, V_{jt}, \mu_{jt}\}$, we then calibrate the parameters in Θ . Appendix C.3 discusses the details of the shooting algorithm.

We first invert the fundamental productivity, $\{\bar{A}_j\}_{j=0}^J$, from the trade balance conditions as specified in equation (18). To do this, we take the w_{jt} and L_{jt} at year 1994, $t = 0$, from the data (see Appendix B.3 for more details), and take the computed $\tau_{ij,0}$ as given. Equation (18) then implies a unique vector of $\{\bar{A}_j\}_{j=0}^J$ (up to a scale) that rationalizes the observed outputs $\{w_{j0}L_{j0}\}$ across locations, including the ROW, in the year 1994. In practice, we normalize $A_{1,0}$, so that the fundamental productivity in location 1, Beijing, to unity. Lastly, note that this step does not require solving the model, as we take the output and trade costs information from the data. However, as the $\tau_{ij,0}$ is a function of Θ^x , we still need to infer productivity at the inner layer for every guess of Θ^x .

We recover the fundamental amenity from the population distribution at $t = 1$ in the data. Different from inverting productivity, where one does not need to solve the model, inferring fundamental amenity requires information on future values of $\{V_{jt}\}$ and thus the entire transition path. We solve the model and then match the endogenous vector $\{L_{j1}\}_{j=1}^J$ to the observed population in the year 1995.¹⁴

The two trade-related parameters, a^f and τ^* , are calibrated as follows. The average cost of freight transportation, a^f , governs the volume of internal trade in China. We follow Ma and Tang (2020) and recover a^f from an average internal-trade-to-GDP ratio of 0.625 between 2000 and 2005, based on the data from the *Investment Climate Survey* conducted by the World Bank. Similarly, τ^* governs the external-trade-to-GDP ratio. We target an average export-to-GDP ratio of China at 24.5 percent between 2000 and 2005, based on the official statistics of China.

The last two parameters, a^p and ψ , govern the migration costs matrix. The average

¹⁴Kleinman et al. (2021) provides a method that backs out fundamental amenity from the observed distribution of population without solving the model in their Online Appendix B.9. To implement their methods, one needs population data from three consecutive years. Such data in China suffer from many quality issues, and therefore we opt to use a computationally heavier method that is more robust to measurement errors in the data. Appendix B.3 discusses these issues.

cost, a^{P} , targets an annual aggregate stay-rate of 98.9 percent between 2000 and 2005. This target comes from the 5-year stay-rate of 94.4 percent as computed using the 2005 *One Percent Population Survey*.¹⁵ The parameter ψ affects the impacts of hukou reform on internal migration. In the data, the aggregate stay rate between 2010 and 2015 had declined substantially to 89 percent one decade later as computed using the 2015 *One Percent Population Survey*. Such a decline in aggregate stay rate comes from the improvements in the transportation networks and the gradual liberalization of the hukou policies. We use ψ to target the average aggregate stay rates in the later decade between 2010 and 2015. Intuitively, to identify ψ , we assume that all the residual changes in aggregate stay rate between the 2000s and the 2010s, conditional on the changes in transportation networks, come from the policy reforms.

4.3.2 Outer Layer: Generalized Method of Moments

The parameters in the outer layer, Θ^{x} , govern the relative costs of different modes of freight and passenger transportation. We estimate these parameters by matching the moments of the mode-specific traffic in the data. From the *City Statistical Yearbooks of China*, we observe the volumes of freight and passenger traffic by prefecture-mode each year. Based on these data, we compute two sets of moments: 1) the average traffic share in each year that identifies $\{a^{m\text{f}}, a^{m\text{p}}\}_{m=1}^M$, and 2) the coefficient of variation of total traffic volume in each year that identifies $\{h^{\text{f}}, h^{\text{p}}\}$.

Moments The first set of moments is the average share of traffic that goes through the road and the rail networks each year across prefectures. The relative traffic shares help to identify $\{a^{m\text{f}}, a^{m\text{p}}\}_{m=1}^M$: conditional on $\{\mathbf{T}_t^{m\text{x}}\}_{m=1}^M$, a mode with higher $a^{m\text{x}}$ is more expensive to travel on, and therefore see less usage in the data. In both the freight and passenger transportation, we normalize the costs of waterway transportation to unity ($a^{3,\text{f}} = a^{3,\text{p}} = 1$) so that the costs of road and rail transportation are identified relative to that of the waterway transportation.

In the model, we compute the average traffic share as follows. Manipulating equation

¹⁵Note that $0.944^{1/5} \approx 0.989$.

(14) yields the volume of sales from j to i via mode m :

$$\sum_{r^m \in \mathcal{R}_{ij}^m} \pi_{ijt,r^m} X_{ij} = \frac{(w_{jt}/A_{jt})^{-\theta} \sum_{K=1}^{\infty} (\mathbf{F}_t^{mf})^K}{\sum_{k=0}^J (w_{kt}/A_{kt})^{-\theta} \tau_{ikt}^{-\theta}} X_{ij} = \frac{(w_{jt}/A_{jt})^{-\theta} b_{ijt}^{mf}}{\sum_{k=0}^J (w_{kt}/A_{kt})^{-\theta} \tau_{ikt}^{-\theta}} X_{ij},$$

From this, the share of sales from j to i via mode m , which we denote as s_{ijt}^{mf} , can be written as:

$$s_{ijt}^{mf} = \frac{b_{ijt}^{mf}}{\sum_{m'=1}^M b_{ijt}^{m'f}},$$

and the share of all the sales from j via mode m , denoted as s_{jt}^{mf} is the weighted average of s_{ijt}^{mf} across destinations excluding ROW, where the weight is X_{ijt}/X_{jt} :

$$s_{jt}^{mf} = \sum_{i=1}^J s_{ijt}^{mf} \frac{X_{ijt}}{X_{jt}}. \quad (21)$$

Lastly, the moments that we match is the vector of average shares across years between 1995 and 2017: $\left\{ \sum_{j=1}^J s_{jt}^{mf} / J \right\}_{t=1}^{23}$. Note that while b_{ijt}^{mf} can be computed without solving the model, the weights $\frac{X_{ijt}}{X_{jt}}$ are endogenous and thus require the knowledge of Υ_t on the transition path.

The share of population flow from j to all the other prefectures via mode m can be computed similarly as:

$$s_{jt}^{m\mathbb{P}} = \sum_{i=1}^J \left(\frac{b_{ijt}^{m\mathbb{P}}}{\sum_{m'=1}^M b_{ijt}^{m'\mathbb{P}}} \right) \frac{L_{ijt}}{L_{jt}}, \quad (22)$$

with the associated moment conditions defined as $\left\{ \sum_{j=1}^J s_{jt}^{m\mathbb{P}} / J \right\}_{t=1}^{23}$. Note that the asymmetry in the migration costs and entry barriers do not affect the mode choice because the mode choice is conditional on a destination.¹⁶

The second set of moments is the coefficient of variation (CV) of the total freight and

¹⁶The passenger volume data in *City Statistical Yearbook* cover both migrants and short-term travelers for business and leisure purposes. By matching $s_{jt}^{m\mathbb{P}}$ to the observed shares, we implicitly assume that there are no systematic differences in the choice of travel modes between migrants and short-term travellers.

passenger traffic across prefectures in each year. These moments help to identify $\{h^f, h^p\}$, the parameters that control the variations in τ_{ijt} and λ_{ijt} conditional on $\{\mathbf{T}_t^{m\chi}\}$.¹⁷ Intuitively, as h^χ declines, the resulting trade or migration costs matrix becomes uniform and less dependent on the underlying geography summarized in $\{\mathbf{T}_t^{m\chi}\}$. Everything else being equal, the resulting variations in trade or migration flows will decline, leading to a smaller CV.

The model counterparts of the CV of total freight and passenger volume in location j , time t are, respectively:

$$CV_{jt}^f = \frac{\text{std}(\{X_{ijt}\}_{i>0 \cap i \neq j})}{\sum_{i>0 \cap i \neq j} X_{ijt} / (J-1)}, \quad CV_{jt}^p = \frac{\text{std}(\{L_{ijt}\}_{i>0 \cap i \neq j})}{\sum_{i>0 \cap i \neq j} L_{ijt} / (J-1)},$$

where $\text{std}(\cdot)$ computes the standard deviation of a vector. The moment conditions are the vector of the average CV across prefecture in each year between 1994 and 2017: $\left\{ \frac{\sum_{j=1}^J CV_{jt}^f}{J} \right\}_{t=1}^{23}$ and $\left\{ \frac{\sum_{j=1}^J CV_{jt}^p}{J} \right\}_{t=1}^{23}$.

Estimation In summary, denote the moment conditions in the data as the vector $\bar{\mathcal{S}}$, and the counter-parts in the model as

$$\mathcal{S}(\Theta^\chi) = \left\{ \frac{\sum_{j=1}^J s_{jt}^{mf}}{J}, \frac{\sum_{j=1}^J s_{jt}^{mp}}{J}, \frac{\sum_{j=1}^J CV_{jt}^f}{J}, \frac{\sum_{j=1}^J CV_{jt}^p}{J} \right\}_{t=1}^{23}.$$

Our estimation strategy is to find Θ^χ to minimize the Euclidean distance between the data and the model moments:

$$\min_{\Theta^\chi} [\bar{\mathcal{S}} - \mathcal{S}(\Theta^\chi)] \mathbf{W} [\bar{\mathcal{S}} - \mathcal{S}(\Theta^\chi)]',$$

where \mathbf{W} is the optimal weighting matrix. In this context, the optimal weighting matrix is the inverse of the variance-covariance matrix of the data moments $\bar{\mathcal{S}}$, computed by bootstrapping the data at the prefecture-mode-year level 500 times. We use the iterative particle swarm

¹⁷We use the ‘‘coefficient of variation’’ instead of ‘‘standard deviation’’ because in the data, the total freight and passenger traffic are recorded in physical units of ‘‘tonnes’’ and ‘‘passengers’’, respectively. The mapping of units between the data and their model counterparts is not clear. In this context, the CV is more suitable because it is scale-free.

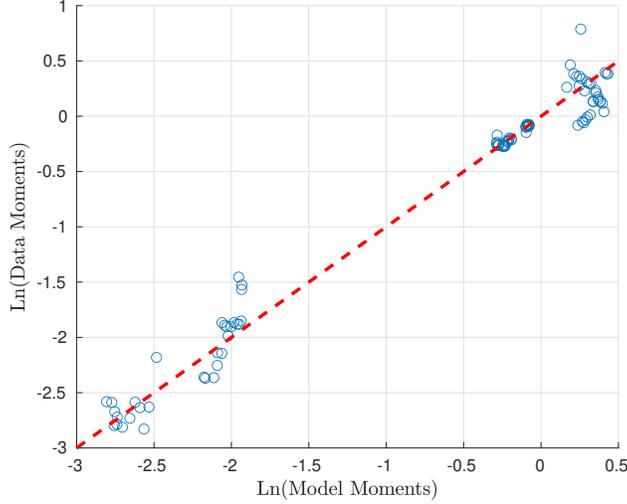


Figure 5: Goodness-of-Fit, GMM

Notes: this figure plots the results of goodness-of-fit of the GMM estimation. The red dashed line is the 45-degree line. The moments on the upper-right corner of the figure are the coefficient-of-variations of freight and passenger traffic in each year. Those on the lower-left corner are the shares of freight and passenger traffic via road and rail network in each year.

optimization algorithm developed in [Ma and Tang \(2020\)](#) to solve the minimization problem and use Gauss-Newton Regression (GNR) to estimate the standard errors.

Results Figure 5 presents the goodness-of-fit graph of the estimation. Overall, the estimation procedure fits the data moments fairly well as most of the data and the model-generated moments cluster around the 45 degree line.

Panel (c) in Table 2 summarizes the estimation results. For both freight and passenger transportation, the road networks are cheaper to use than the rail system, as reflected in $a^{1,\chi} < a^{2,\chi}, \forall \chi$. This result is driven by the fact that most freight and passenger traffic goes through the road system. For example, 79 percent goes through the road system on average for freight transportation. Only 14 percent of the traffic goes through the rail system. Passenger transportation is even more reliant on the road system with 91 percent of the traffic, leaving only 7 percent on the rail system.

5 Quantitative Results

In this section, we present the quantitative results based on the estimated parameters from the previous section. We start by assessing the welfare and the distributional impacts of the transportation network improvement over the entire period; we then decompose the impact into trade v.s. migration channels to highlight the mechanisms through which transportation networks affect welfare and spatial inequality.

We evaluate the impacts of transportation networks by comparing two sets of simulations, “baseline” and “no-change”. In both simulations, we start with the initial population distribution in the year 1995 and solve the transition path 50 years forward until 2044. In the “baseline” simulation, the τ_{ijt} and λ_{ijt} matrices are based on the existing transportation networks and policy barriers in each year between 1995 and 2017 and are fixed to their levels in 2017 for all the subsequent years. In the “no-change” counterfactual, we fix both the τ_{ijt} and the λ_{ijt} matrices to their levels in the initial year 1995. In other words, the baseline case simulates the economy that converges to a steady state as defined by the transportation networks in 2017, and the “no-change” case is the transition path towards a steady state implied by the initial conditions in 1995. Comparing the two sets of results reveals the impacts from the expansion of the transportation networks.

To further decompose the impacts of the transportation networks through trade and migration, we also simulate two other counterfactual transition paths, “ τ -only” and “ λ -only”. As the names suggest, in the “ τ -only” case, we fix the values of λ_{ijt} to the 1995 levels and allow τ_{ijt} to evolve in the same way as in the baseline. In the “ λ -only” case, we do the reverse and only allow λ_{ijt} to change over time. The results of these exercises are summarized in Figures 6 to 8. We first briefly discuss the aggregate impacts of transportation networks and then delve into the distributional impacts.

5.1 Aggregate and Dynamic Impacts

Figure 6 summarizes the impacts of transportation networks on aggregate output. In this figure, we plot the transition paths of aggregate output in the four counterfactual simulations and normalize the initial levels in 1995 to 1. We first highlight the long-run impacts in the

steady state, and then study the transition paths.

Steady State The growth of transportation networks significantly improved aggregate output in the long run. In the “no-change” case, the aggregate economy grows by 10.8 percent, a benchmark against which we compare the other counterfactual cases. The aggregate economy in the “no-change” simulation grows because the initial year is on a transition path to a long-run steady state. Along this transition path, workers migrate to more productive prefectures, attracted by the higher real wage, and thus increase aggregate output. In the baseline case, the aggregate output grows by 17.3 percent over the same period, which is $17.3/10.8 - 1 \approx 60$ percent higher.¹⁸ The aggregate return to transportation comes from both trade and migration. From the trade channel, better infrastructure allows for higher gains from trade. At the same time, lower migration frictions facilitate population flows toward the more productive locations.

The long-run contributions of trade and migration liberalization are roughly equal. The two dashed lines in Figure 6 report the transition paths of the τ -only and the λ -only cases. Towards the end of the simulations, the reductions in τ alone lead to a 13.7 percent growth, or $13.7/10.8 - 1 \approx 27$ percent faster than the no-change case; those in λ lead to slightly higher output growth at 14.9 percent, or $14.9/10.8 - 1 \approx 38$ percent faster. The net impacts of trade and migration add up to $27 + 38 = 65$ percent faster growth, which is only slightly higher than the gain in the baseline case (60 percent). This result suggests a positive but mild interaction between the two channels.

Dynamics While the return to trade costs reduction is high in the short-run and declines in the long run, those to the migration costs reduction could be negative in the short run but grows substantially over time. For example, before 2015, while the aggregate output under τ -only case is higher than the no-change scenario, those under the λ -only simulation are often *worse* than the no-change case. In other words, the short-run return to migration liberalization is negative for the first 20 years of the transition path. By 2017, almost all the

¹⁸The impact on aggregate economic growth is modest relative to the ten-fold economic growth in the data over the same period. The model is not designed to explain the first moments as we have abstracted away from many factors that lead to higher aggregate total factor productivity (TFP), such as technology improvements, physical capital formation, and human capital investments.

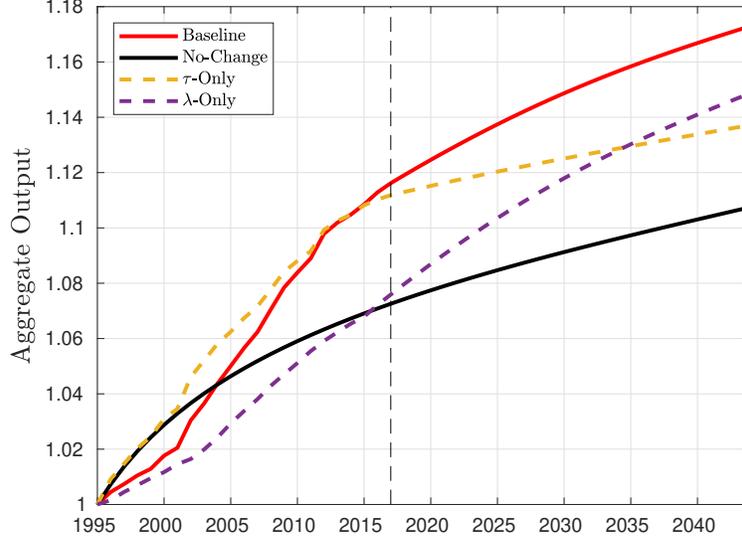


Figure 6: The Aggregate and Dynamic Impacts of Transportation Networks

Notes: this figure plots the transition paths of total output in China in four simulations. “Baseline” refers to the case in which both τ and λ matrices evolve based on the actual data between 1995 and 2017, and then fixed to their 2017 levels in subsequent years, respectively. “No-Change” is the case in which both τ and λ are fixed to their 1995 levels. “ τ -only” is the case in which only λ is fixed at the 1995 level, and “ λ -only” is the case in which only τ is fixed at the 1995 level. In all the cases, the initial output level is normalized to 1 in the year 1995. The vertical dashed line indicates 2017, the last year we have data on transportation networks.

returns to transportation networks come from the improvements in τ , and only a negligible fraction comes from the improvements in λ . However, the return to migration liberalization starts to dominate in the long run. By the end of the simulation in 2044, the return to migration is already higher trade, as noted earlier in this part.

The immediacy of return to trade liberalization is straightforward because the trade decisions are static in the model, as seen in equation (18). In other words, the firms and individuals immediately respond to the changes in the trade network without any forward-looking concerns. In the long run, the return to trade liberalization wanes because of the equalizing effect of trade that we will discuss in detail later. In short, better market access benefits remote locations as it substantially reduces their price index. As a result, lower τ tends to keep the population in remote locations. However, as these remote locations often lack fundamental productivity, in the long run, the equalizing effects of trade are a double-edged sword that leads to slower growth in aggregate output.

The delayed return to migration liberalization is because the migration decisions are

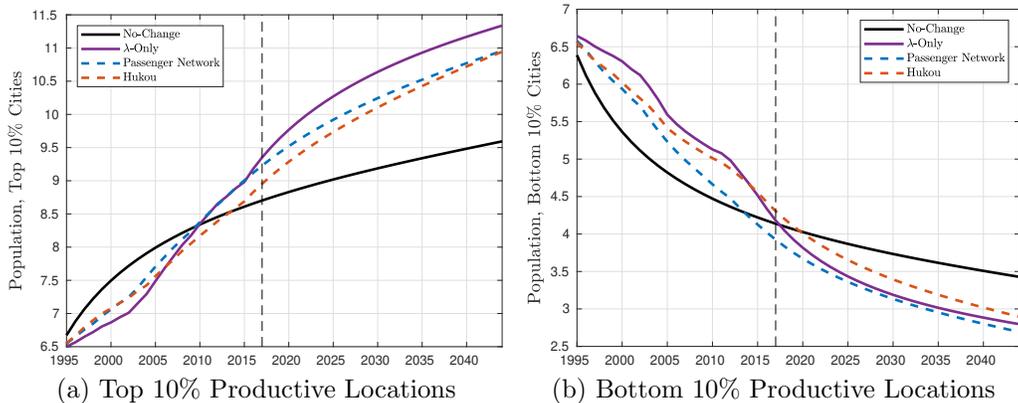


Figure 7: The Dynamic Response to Migration Liberalization

Notes: this figure plots the transition paths of population flows in four simulations. “No-Change” is the case in which both τ and λ are fixed to their 1995 levels. “ λ -only” is the case in which τ is fixed at the 1995 level, and λ changes over time. “Passenger Network” is the case in which only \mathbf{T}_t^{mP} changes over time, and “Hukou” is the case in which only \bar{d}_{jt} changes over time. The vertical dashed line indicates 2017, the last year we have data on transportation networks.

forward-looking. If individuals expect a better passenger travel network and lower hukou restrictions in the future, they might delay the intended move to the more productive prefectures. Figure 7 confirms the dynamic impacts on population movements by plotting the evolution of the population in the most and least productive prefectures. As individuals expect lower migration costs in the future, the migration towards to most productive prefectures is delayed, as shown in Panel (a). Instead, people choose to prolong their stay in less productive locations, as shown in Panel (b). Both forces lead to lower economic output in the short run. In the long run, however, the return to migration liberalization is substantial and long-lasting, as it facilitates population movements toward the more productive locations. Figure 7 also highlights that both the passenger travel network (\mathbf{T}_t^{mP}) and the policy barriers (\bar{d}_{jt}) have similar dynamic impacts on population movements.

5.2 Distributional Impacts

In the last part, we discuss the distributional impacts of transportation networks. We measure spatial inequality using the convergence of real wage, commonly known as “ β -convergence” in the macroeconomics literature following Barro and Sala-i-Martin (1992). In particular, we regress the growth rate of a variable of interest between 1995 and 2044

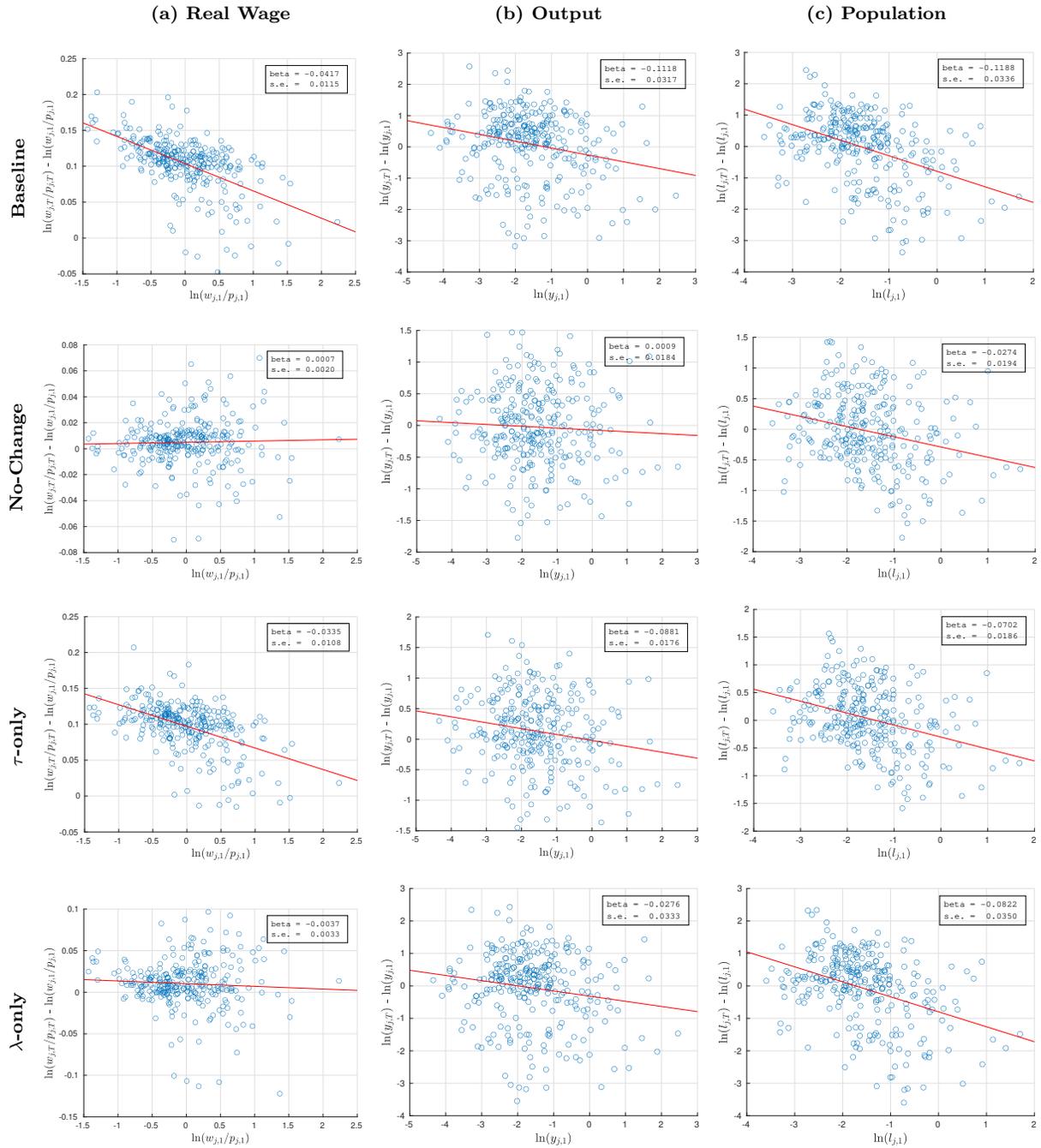


Figure 8: Transportation Networks On β -Convergence

Notes: The figures present the impacts of transportation network improvements on β -convergence between 1995 and 2017 of real wage, output, and population. The β -convergence is the coefficient of regressing the growth rate of a variable between 1995 and 2017 against the logarithm of its initial levels in 1995. Each dot represent a prefecture.

against initial level in 1995 in each prefecture. A negative coefficient suggests that the richer prefectures in 1995 grew slower than the poorer ones, and therefore the spatial inequality must have declined over time. Figure 8 summarizes our findings.

The expansion of transportation networks has significantly reduced spatial inequality. In the baseline case, as shown in the first row of Figure 8, the distribution of economic activity measured by real wage, output, and population, exhibits strong convergence as the β parameters are significantly negative at -0.0417 , -0.112 , and -0.119 , respectively. In stark comparison, without the changes in the transportation networks, spatial inequality would have stayed the same as in 1995. As shown in the second row of Figure 8, the convergence parameters for all three variables in the “no-change” case are not significantly different from zero. In other words, these results show that almost all the convergence in economic activities in China comes from the changes in transportation networks. Our results stand in contrast with those from the U.S., as shown in Kleinman et al. (2021). In that paper, the authors showed that the initial conditions in the U.S. during the 1960s predicted regional convergence in the next several decades without any future changes in the location fundamentals and transportation networks. The contrast between our results highlights that what drives regional convergence is perhaps country-specific, depending on many factors such as the position on the transition path, the long-run steady-state, and the underlying changes in transportation infrastructure and policy.

In theory, transportation networks bring about two counter-acting forces that could shape spatial inequality in either direction. On the one hand, lower trade friction facilitates inter-city trade and reduces spatial inequality. The smaller and remote prefectures benefit more from trade liberalization as the improved market access to the large cities significantly reduces their price index. On the other hand, better infrastructure also reduces migration frictions, thus increasing spatial inequality. The expansion of the passenger travel networks allows people from unproductive prefectures to migrate to productive ones. The agglomeration externality then implies that a higher population in the prefectures with higher fundamental productivity further widens the existing productivity differences, leading to higher spatial inequality. The overall impacts of transportation networks on spatial inequality are quantitative, as they depend on underlying parameters in the model, the location fundamentals,

and the actual changes in the networks.

In the case of China, the reduction in internal trade frictions is the main driver of regional convergence in real wage and output. To study the source of regional convergence, we again decompose the impacts into the “ τ -only” and the “ λ -only” cases and report the results in the third and fourth rows of Figure 8. If better infrastructure only facilitates inter-city trade, then the convergence parameter of real wage stands at -0.0335 , which is around $0.0335/0.0417 \approx 80$ percent of the convergence in the baseline case. In comparison, in the λ -only case, the contribution is one order of magnitude smaller at $0.0037/0.0417 \approx 9$ percent of the baseline case, and it is not significantly different from zero. The convergence of output shown in the second column of Figure 8 conveys a similar message. In that case, the reductions in τ explain around $0.0881/0.112 \approx 79$ percent of the convergence, while those in λ are responsible for $0.0276/0.112 \approx 25$ percent but not significantly different from zero. The role of internal trade as an equalizer is not a surprise, as better market access often benefits poorer and remote locations disproportionately, as explained above.

Both τ and λ are equally important in the convergence of population, as shown in the last column of Figure 8. In this case, τ explains around $0.0702/0.119 \approx 59$ percent while λ explains around $0.0822/0.1188 \approx 69$ percent of the baseline convergence in population. The contribution of τ comes naturally from the equalizing effects of trade. In the λ -only case, the regional inequality in population also declines because individuals flow towards richer prefectures, which are not necessarily the populous ones. In our estimation, the correlation between the initial population and $\{\bar{A}_j\}$ is positive but low at only 0.06.

6 Conclusion

This project studies the aggregate and the distributional impact of transportation networks in China. To do this, we first provide a comprehensive panel dataset on the expansion of the transportation networks in China between 1994 and 2017. Compared to the existing transportation measures, our dataset consistently measures the quality of connection across time and space. Based on this dataset, we evaluate the impacts of transportation networks via a dynamic spatial general equilibrium model that features forward-looking migration

decisions, mode and route choices for freight and passenger transportation, and intercity and international trade. We convey several main messages.

We show that the investments in infrastructure have significantly contributed to the economic growth in China, and the returns of trade and migration liberalizations show rich dynamics. While the return to trade costs reductions are high in the short-run, it declines in the long run as better market access dissuades workers from migrating to productive locations. On the other hand, the short-run return to migration liberalization could be harmful, as forward-looking individuals delay migration in anticipation of better road access and lower policy restrictions in the future. However, in the long run, migration liberalization pays off as it facilitates population flow towards more productive locations.

We also show that in the case of China, the expansion of transportation networks is the major, if not the sole, driver of regional convergence in real wage, output, and population. The equalizing effect of better connectivity mainly comes from the facilitation of internal trade, as it allows better market access to remote locations.

References

- Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf**, “The economics of density: Evidence from the Berlin Wall,” *Econometrica*, 2015, 83 (6), 2127–2189.
- Alder, Simon, Michael Zheng Song, and Zhitao Zhu**, “Unequal Returns to China’s Intercity Road Network,” 2021. working paper.
- Allen, Treb and Costas Arkolakis**, “Trade and the Topography of the Spatial Economy,” *The Quarterly Journal of Economics*, 2014, 1085, 1139.
- and –, “The Welfare Effects of Transportation Infrastructure Improvements,” *The Review of Economic Studies*, 02 2022.
- Anderson, James E. and Eric van Wincoop**, “Trade Costs,” *Journal of Economic Literature*, September 2004, 42 (3), 691–751.

- Artuc, Erhan, Shubham Chaudhuri, and John McLaren**, “Trade Shocks and Labor Adjustment: A Structural Empirical Approach,” *The American Economic Review*, 2010, *100* (3), 1008–1045.
- Banerjee, Abhijit, Esther Duflo, and Nancy Qian**, “On the Road: Access to Transportation Infrastructure and Economic Growth in China,” Working Paper 17897, National Bureau of Economic Research March 2012.
- Barro, Robert J. and Xavier Sala-i-Martin**, “Convergence,” *Journal of Political Economy*, 1992, *100* (2), 223–251.
- Baum-Snow, Nathaniel**, “Did Highways Cause Suburbanization?*,” *The Quarterly Journal of Economics*, 05 2007, *122* (2), 775–805.
- , **J. Vernon Henderson, Matthew A. Turner, Qinghua Zhang, and Loren Brandt**, “Does investment in national highways help or hurt hinterland city growth?,” *Journal of Urban Economics*, 2020.
- Caliendo, Lorenzo, Fernando Parro, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte**, “The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy,” *The Review of Economic Studies*, 2018, *85* (4), 2042–2096.
- , **Maximiliano Dvorkin, and Fernando Parro**, “Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock,” *Econometrica*, 2019, pp. 741–835.
- Donaldson, Dave**, “Railroads of the Raj: Estimating the Impact of Transportation Infrastructure,” *American Economic Review*, April 2018, *108* (4-5), 899–934.
- **and Richard Hornbeck**, “Railroads and American Economic Growth: A “Market Access” Approach *,” *The Quarterly Journal of Economics*, 02 2016, *131* (2), 799–858.
- Eaton, Jonathan and Samuel Kortum**, “Technology, Geography, and Trade,” *Econometrica*, Sep. 2002, *70* (5), 1741–1779.

- Faber, Benjamin**, “Trade Integration, Market Size, and Industrialization: Evidence from China’s National Trunk Highway System,” *The Review of Economic Studies*, 03 2014, 81 (3), 1046–1070.
- Fan, Jingting**, “Internal Geography, Labor Mobility, and the Distributional Impacts of Trade,” *American Economic Journal: Macroeconomics*, 2019, 11 (3), 252–88.
- , **Yi Lu, and Wenlan Luo**, “Valuing Domestic Transport Infrastructure: A View from the Route Choice of Exporters,” *The Review of Economics and Statistics*, 08 2021, pp. 1–46.
- Gurara, Daniel, Vladimir Klyuev, Nkunde Mwase, and Andrea F Presbitero**, “Trends and Challenges in Infrastructure Investment in Developing Countries,” *International Development Policy—Revue internationale de politique de développement*, 2018, 10 (10.1).
- Huang, Sixiang and Dong Mu**, “Discussion on the development strategy of China’s multimodal transport stations,” *Advances in Intelligent Systems Research*, 2018, 161, 256–262.
- Kapos, Valerie, J Rhind, Mary Edwards, MF Price, C Ravilious et al.**, “Developing a map of the world’s mountain forests.” *Forests in sustainable mountain development: a state of knowledge report for 2000. Task Force on Forests in Sustainable Mountain Development.*, 2000, pp. 4–19.
- Kleinman, Benny, Ernest Liu, and Stephen J Redding**, “Dynamic Spatial General Equilibrium,” Working Paper 29101, National Bureau of Economic Research July 2021.
- Lin, Yatang**, “Travel costs and urban specialization patterns: Evidence from China’s high speed railway system,” *Journal of Urban Economics*, 2017, 98, 98 – 123. Urbanization in Developing Countries: Past and Present.
- Ma, Lin and Yang Tang**, “Geography, trade, and internal migration in China,” *Journal of Urban Economics*, 2020, 115, 103181.

- Ma, Liqian**, *The Chronicle of Railway Construction in China (in Chinese)*, China Railway Press, 1983.
- Qin, Yu**, “‘No county left behind?’ The distributional impact of high-speed rail upgrades in China,” *Journal of Economic Geography*, 06 2016, *17* (3), 489–520.
- Redding, Stephen J.**, “Goods trade, factor mobility and welfare,” *Journal of International Economics*, 2016, *101*, 148–167.
- **and Matthew A. Turner**, “Chapter 20 - Transportation Costs and the Spatial Organization of Economic Activity,” in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5 of *Handbook of Regional and Urban Economics*, Elsevier, 2015, pp. 1339 – 1398.
- Rozenberg, Julie and Marianne Fay**, *Beyond the gap: How countries can afford the infrastructure they need while protecting the planet*, The World Bank, 2019.
- Tombe, Trevor and Xiaodong Zhu**, “Trade, Migration, and Productivity: A Quantitative Analysis of China,” *American Economic Review*, May 2019, *109* (5), 1843–72.
- Xu, Mingzhi**, “Riding on the new silk road: Quantifying the welfare gains from high-speed railways,” 2017. working paper.

A Tables and Figures

Table A.1: The Physical Maps

Year	National Maps Publisher	Scale	Projection	Provincial Map Publisher
1994	Sino Maps	1:6 million	Albers, 25N, 47E	Sino Maps
1995	N/A			Sino Maps
1996	Sino Maps	1:4.5 million	Albers, 25N, 47E	Global Maps
1997	N/A			Global Maps
1998	N/A			Xi'an Maps
1999	N/A			Xi'an Maps
2000	Sino Maps	1:6 million	Albers	Xi'an Maps
2001	N/A			Xi'an Maps
2002	Sino Maps	1:4.5 million	Albers, 25N, 47E	Xi'an Maps
2003	Sino Maps	1:6 million	Albers, 25N, 47E	Xi'an Maps
2004	N/A			Xi'an Maps
2005	N/A			Xi'an Maps
2006	N/A			Hunan Maps
2007	Guangdong Maps	1:6 million	Lambert, 24N, 46N, 110E	Dizhi
2008	N/A			Dizhi
2009	Sino Maps	1:4.5 million	Albers, 25N, 47E	Renmin Jiaotong
2010	N/A			Dizhi
2011	N/A			Dizhi
2012	Sino Maps	1:4.5 million	Albers, 25N, 47E	Dizhi
2013	Sino Maps	1:4.6 million	Albers, 25N, 47E	Renmin Jiaotong
2014	N/A			Sino Maps
2015	N/A			Sino Maps
2016	N/A			Sino Maps
2017	Sino Maps	1:6 million	Albers	Sino Maps

Non-Map References	
Years	Title
1986-2017	Transportation Yearbooks of China
1994-2017	Railway Yearbooks of China
1881-1981	The Chronicle of Railway Construction in China (Ma, 1983)

Notes: this table presents the basic information on the physical maps in our collection. The Hunan, Xi'an, and Guangdong Maps are regional publishers, while the others are national. Both the Albers and the Lambert projections are conic projections; the Albers projection is an equal-area projection while the Lambert projection is conformal. The coordinates following the projections are the reference longitude and latitudes. The coordinate (25N, 47E) is commonly used for Chinese maps as it centers around Henan, the geographical center of China.

Table A.2: The Mapping Between Road Classification and Map Legends

Year	First Rate Road	Highway	Railroad	High Speed Rail
1994	Zhong Yao Gong Lu	Gao Su Gong Lu	Dian Qi Hua Tie Lu	-
1994			Shuang Gui Tie Lu	-
1994			Dan Gui Tie Lu	-
1995	Zhu Yao Gan Xian Gong Lu	Gao Su Gong Lu	Dian Qi Hua Tie Lu	-
1995			Fu Xian Tie Lu	-
1995			Dan Xian Tie Lu	-
1995			Zhai Gui Tie Lu	-
1996	Zhong Yao Gong Lu	Gao Su Gong Lu	Dian Qi Hua Tie Lu	-
1996			Shuang Gui Tie Lu	-
1996			Dan Gui Tie Lu	-
1997	Zhu Yao Gong Lu	Gao Deng Ji Gong Lu	Tie Lu	-
2000	Guo Dao	Gao Su Gong Lu	Tie Lu	-
2002	Guo Dao	Gao Deng Ji Gong Lu	Tie Lu	-
2003	Guo Dao	Gao Deng Ji Gong Lu	Tie Lu	-
2007	Guo Dao	Gao Su Gong Lu	Tie Lu	-
2008	Guo Dao	Gao Su Gong Lu	Tie Lu	-
2009	Guo Dao	Gao Deng Ji Gong Lu	Tie Lu	-
2012	Guo Dao	Gao Deng Ji Gong Lu	Tie Lu	-
2013	Guo Dao	Gao Su Gong Lu	Tie Lu	Gao Su Tie Lu
2017	Guo Dao	Gao Deng Ji Gong Lu	Tie Lu	-

Notes: this table shows the correspondence between map legends and the classification of transportation modes used in this paper. We directly report the Pinyin of the legends.

Table A.3: List of Port Cities

Tianjin	Shanghai	Fuzhou	Qingdao	Shenzhen	Shantou	Zhanjiang
Dalian	Ningbo	Xiamen	Guangzhou	Zhuhai	Jiangmen	Haikou

Notes: this table lists the 14 prefectures that 1) imports and exports from the international markets in the Chinese Customs database and 2) are on the coast.

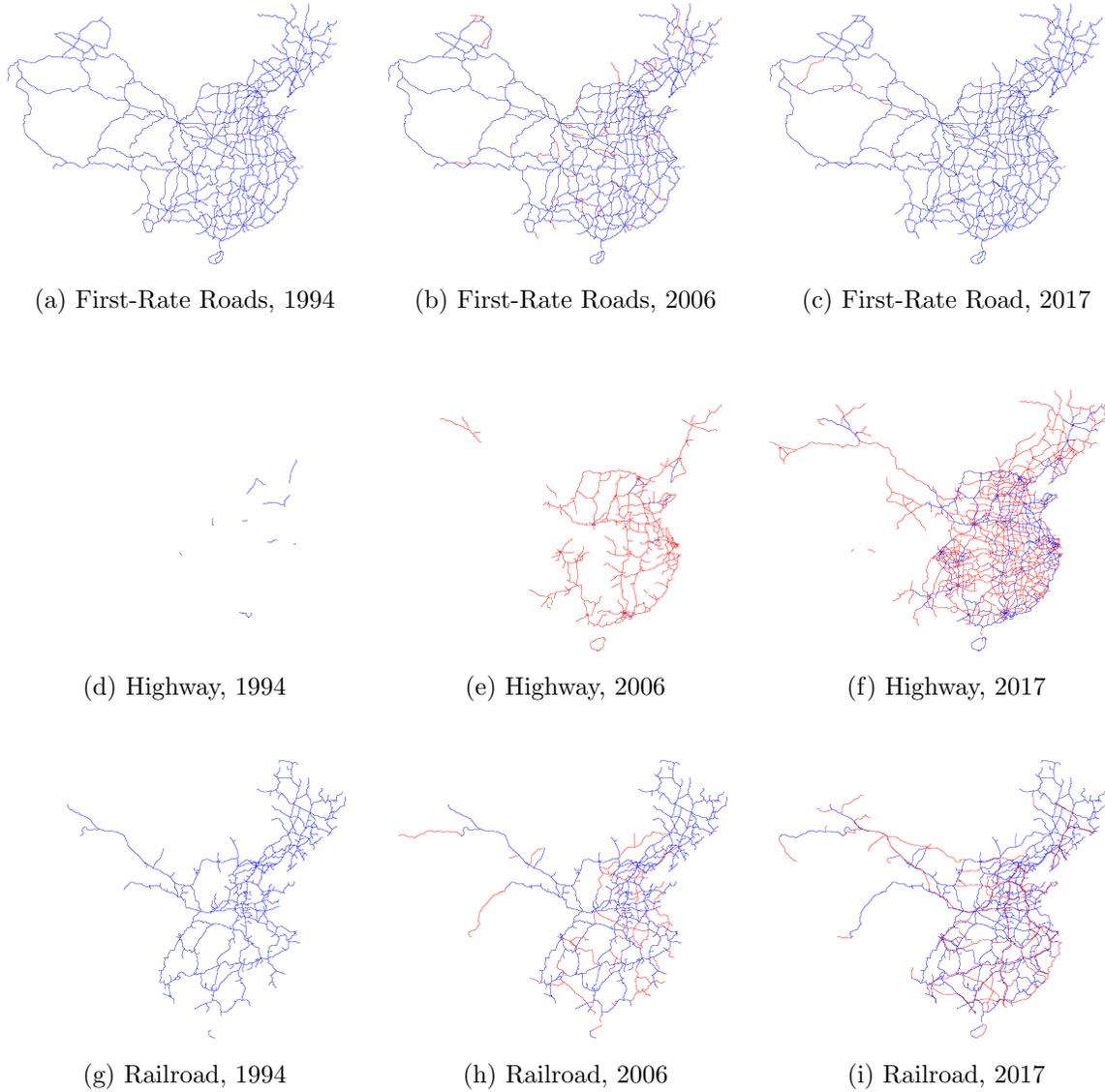


Figure A.1: Changes in Transportation Networks, 1994-2017

Notes: this figure presents the evolution of transportation networks between 1994 and 2017 in several selected years. The “railroad” network includes all the freight-only, passenger-only, and mixed-use railroads. Blue lines indicate existing connection by year t , and the red lines indicate new construction between the year in the previous figure and the current figure (e.g., between 1994 and 2006 in Panel b).

B Dataset Construction

B.1 Measuring Connectivity

The physical maps that are digitized in this project are summarized in Table A.1 and A.2. We use standard procedures to digitize these maps: after scanning the maps, we extract the modes of transportation by color identification and then geo-reference each map by inverting the projection methods. Despite the differences in projection methods and scale, the geo-referenced maps are comparable across the years at the pixel level. The dimension of the digitized national maps is 12669 pixels in width and 8829 pixels in height, which implies that each pixel represents around 500 square meters in the real world.

The transportation networks in the color-identified maps usually have paths that are 10 to 20 pixels in width. Given that each pixel corresponds to around 500 square meters in the real world, we shrink the color-identified maps down to one-pixel paths using the skeletonization algorithm. We use eight-connectivity at the pixel level, which means that each pixel in the network is considered to be connected to all eight of its neighbors, including the diagonal neighbors. We then break the entire skeleton network into “segments”, defined as the set of pixels between the branch and endpoints of the graph. See Figure B.1 for an illustration. The above procedure is applied to all modes of transportation and in all years.

To ensure that each segment is consistently represented over time, we compare the segments in nodal year t_{i+1} to all the segments in nodal year t_i to determine if the segment already existed in year t_i , or it was constructed between t_i and t_{i+1} . To classify the roads, we first denote segment k in year t_{i+1} as $g_{t_{i+1}}^k$, and then use the following procedure:

1. If there exists a segment in year t_i , denoted as $g_{t_i}^{k'}$ such that $g_{t_{i+1}}^k \subseteq g_{t_i}^{k'}$, then segment k already existed in t_i .
2. If no such segment exists in the previous step, we then manually compare $g_{t_{i+1}}^k$ to the closest segment in t_i^k , and determine if $g_{t_{i+1}}^k$ already existed in t_i . To determine the closest segment, we first compute pair-wise distance between the all the pixels in two paths, and take the simple average. To determine if the segment already existed, we visually compare the segments in the national maps and also refer to the provincial

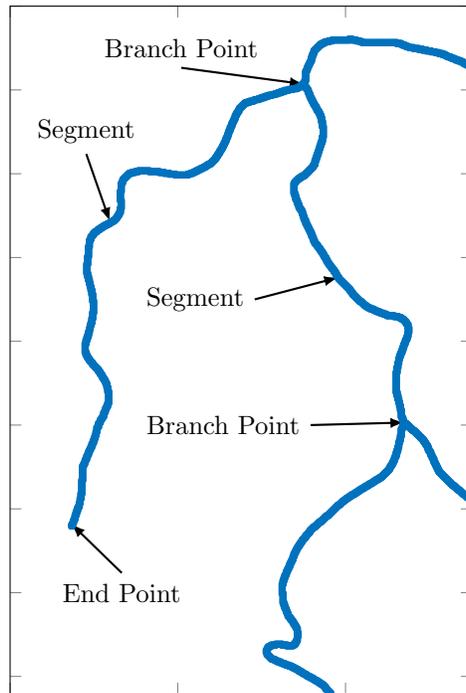


Figure B.1: Definition of Segments, End Points, and Branch Points

Notes: this figure illustrates the definition of “segments” in our dataset. The blue lines indicate a transportation network. A “segment” is the set of pixels in between any two branch points or end points of the network.

maps, yearbooks, and chronicles.

3. For newly constructed segments, we then refine the segments to avoid overlapped pixels. As shown in the example in Figure B.2 and explained in details later, a segment in t_{i+1} might contain pixels that are already constructed in t_i . Therefore, we refine the segment $g_{t_{i+1}}^k$ to drop all the pixels that already exists in $\cup_{k'} g_{t_i}^{k'}$.
4. To determine the year of construction, we first refer to the yearbooks and the chronicles because they are official publications with the exact date of construction and the relevant contexts. The *Transportation Yearbooks* mainly cover the highway and significant railway construction, while *Railway Yearbooks* and the *Chronicles of Railway Construction* together provide a complete picture of railway construction. If a segment does not show up in any of the yearbooks or chronicles, we then rely on visual identification from the provincial map collections to determine the year of construction. The first-rate roads often fall into this category as they are not deemed important enough to be recorded in the yearbooks.
5. For railway construction, yearbooks and chronicles could record up to four milestone dates in chronicle order: construction finished (Tong Che), trial-run (Shi Yun Ying), in-business (Yun Ying), and record-closure (Xiao Hao). Unfortunately, many railways do not have a complete record of all four dates, and therefore we do not have the liberty to designate one type of date as the “date of construction”. As any of the milestone dates could be interpreted as the railway’s completion, we use the first mention of any of the four dates as the date of construction.

Example In this part, we provide an example of the identification of a newly constructed highway, as shown in Figure B.2. In this figure, the thin red line is a segment in the second nodal year, $t_2 = 1996$, identified by breaking up the graph in 1996 between all the branch points and endpoints. The thick blue line is the nearest segment in the previous nodal year, $t_1 = 1994$. The figure shows that the two lines overlap, as each is identified separately relative to the network in their respective nodal years. The map shows that the highway in question is the Chengyu Highway, which connects Chengdu and Chongqing in Southwest China. From

the yearbooks, the blue line is the early stage of the highway that already existed by 1994. The newly constructed red line is the phase-two construction that was finished in 1995, as recorded in the *Transportation Yearbook of China, 1996*, page 438, translated and quoted here for reference:

“(The province of Sichuan) had made breakthroughs in highway construction. Beating the original plan, the Chengyu Highway started a trial run on July 1st and commenced operation on September 25th (1995). The highway, measuring 340.2 km, constructed in 5 years with a cost of 4 billion RMB, was the very first highway in Sichuan.”



Figure B.2: Identification of New Construction

Notes: this is an example of a newly constructed highway. In the figure, the thin red line is a segment in $t_2 = 1996$, and the thick blue line is the nearest segment in the previous nodal year, $t_1 = 1994$. Based on this and the records in the yearbooks, we determine that the differences between the red and the blue lines form a new segment that was constructed between 1994 and 1996. The highway in question is the Chengyu Highway which connects Chengdu and Chongqing. The blue segment is the phase-one construction that was already finished by 1994, and the newly constructed red line is the phase-two construction that was finished in 1995, as recorded in the *Transportation Yearbook of China, 1996*, page 438.

Lastly, we refine the red segment to drop the overlapped pixels with the existing blue segment and assign the construction year of 1995 to the re-defined segment.

B.2 Measuring Quality

Road To measure the quality of roads across time and space, we rely on the publications from the Ministry of Transportation from China: the *Technical Standard of Highway Engineering*. We use the following four revisions:

1. 1988 (JTJ01-88)
2. 1997 (JTJ01-97)
3. 2003 (JTG B01-2003)
4. 2014 (JTG B01-2014)

We apply the design speed from the prevailing design codes at the time to determine the quality of highways and the first-rate roads. For example, JTJ01-88 regulates the highways constructed in 1995, and JTJ01-97 should apply to those constructed in 1999. As highway construction in China started in 1988, the four revisions covered all the highways constructed in China. The case of first-rated roads is tricky. While we can determine the applicable design codes for the first-rate roads built after 1994, it is impossible to determine the years of construction of those that already existed in 1994 based on our data. We apply the 1988 revision to all the existing first-rated roads in 1994 when measuring quality. This exercise assumes that by assigning a particular road as “the first-rate” in 1994, that road must have a quality that is close to those defined by the prevailing standards at the moment, the JTJ01-88 revision; lower quality roads would have been assigned a lower rate on the national maps.

The information on design speed comes from the *Technical Standard of Highway Engineering*. In the 1988 revision, the design speed, written as “Ji Suan Xing Che Su Du” as in Pinyin, depends on the terrain as stipulated in Chapter 2.0.2. The dependence on terrain is particularly emphasized for the construction of highways due to the difficulties and costs in construction. In the mountainous regions, the default design speed is 80km/h, and the 60km/h speed is reserved for “challenging segments”. As no more detail is provided in defining the conditions for using the 60km/h design speed, we choose to use 80km/h for all the highways in the mountain areas constructed under the 1988 revision. The design speed

for the first-rate roads are clear: 100km/h in plains and low rolling hills and 60km/h in the high hills and mountains.

The major change in the 1997 revision is to remove the dependence on terrain for highway construction. This change is explicitly stated in the notes as “the design speed of highway is no longer linked to the underlying terrain, ..., under normal circumstances, the design speed should be 120km/h”. Under limited conditions, the design speed can be lowered to 100 or 80 km/h. In “particularly difficult segments”, 60km/h is still admissible. For this reason, we use a design speed of 120km for all terrains except for the mountains, for which we still use the 80km/h design speed. The construction of the first-rate roads is still reliant on the underlying terrain, and there is no change in the design speeds.

The next revision in 2003 focused on first-rate roads. In high hills and mountains, the design speed of this class of roads increased from 60km/h to 80km/h, the same as the highways. In the highway construction, the 60km/h design is explicitly discouraged, as the Ministry learned that it is challenging to upgrade a highway with a 60km/h design to accommodate the ever-increasing traffic volume. The 2003 revision still allows a 60km/h design for highways, but it stipulated that such segments cannot be longer than 15km. The 2003 revision also changed the Chinese translation of “design speed” from “Ji Suan Xing Che Su Du” to a more direct translation of “She Ji Su Du” to be comparable to the international standards.

The design speed of the highways and the first-rate roads did not change in the latest 2014 revision. The latest revision stated that the 2003 revision was already well-crafted and was widely tested in practice. For consistency, the Ministry no longer saw a need to revise the design speeds further.

Railroad The evolution of railroad engineering design codes is much more convoluted than that of the highways. Several strands of codes exist, and the classification of rates also changes over time. Table [B.1](#) summarizes the mapping between railroad codes and the rates railroads that they cover. In the rest of this part, we discuss several specific issues related to railroad design codes.

First, unlike highway engineering, the design codes for railway engineering do not map

Table B.1: The Mapping Between Railroad Rates and the Codes of Railroad Design

Codes: Revision:	<i>Code for Design of Railway Line</i>				CDRL (III,IV) ^a 2012	CDSRLIF ^b 1987
	1985	1999	2006	2017		
Doc.Number:	GBJ90-85	GB50090-99	GB50090-2006	TB10098-2017	GB50012-2012	GBJ12-87
National I:	■	■	■	■	□	□
National II:	■	■	■	■	□	□
National III:	■	■	□	□	■	□
National IV:	□	□	□	□	■	□
Local I:	□	□	□	□	■	□
Local II:	□	□	□	□	■	□
Local III:	□	□	□	□	■	□
Industrial I:	□	□	□	□	□	■
Industrial II:	□	□	□	□	□	■
Industrial III:	□	□	□	□	□	■

Notes: this table lists the mapping between the railway design codes with various rates of railroads. ■ indicates that the design code covers the railroad rate in question, and □ indicates otherwise.

^aCode for Design of III and IV Rated Railway Line.

^bCode for Design of Standard Railway Line for Industrial Firms.

speed to terrain in every revision. The 1999 revision (GB50090-99) was the only one that explicitly mapped design speed and terrain in Chapter 1.0.5, Table 1.0.5-2 for National I and II railroads. Other revisions still emphasize design speed’s dependency on terrain but do not explicitly provide a mapping between the two. For example, in the 2006 revision, Chapter 1.0.5 defines the speed of National I railroads at 160km/h, 140km/h, and 120km/h, depending on terrain, but does not provide any further indication.

We use the mapping defined in the 1999 revision as a reference and infer the mapping between speed and terrain in the other revisions to address this issue. In principle, whenever the design code allows for several categories of design speed within a rate, we assign the highest design speed to the plains, followed by LRH, hills, and mountains. If the design code allows for four-speed categories, such as National IV in the 2012 revision, we map the four speeds to the four types of terrain. In the case of three-speed categories, such as National III in the 2012 revision, we assign the same speed to “hills” and “mountains”, following the mapping of National I in the 1999 revision. In the case of two-speed categories, we assign the same speed to LRH, hills, and mountains, following the mapping of National II in the 1999 revision. For example, the 2006 revision stipulated three categories of speed for National I railroads at 160, 140, and 120 km/h. Based on this, we assume the design speed on plains

to be 160km/h, LRH to be 140km/h, and hills and mountains to be 120km/h.

Second, different from the road rating system, the rates of railroads had changed significantly over time. In earlier revisions, the Ministry of Railroads rates railroads by National I, II, and III, followed by Local I, II, and III ratings. The “local” rated railroads were typically constructed with a lower quality standard. However, starting from the 2012 revision, local ratings were discontinued, and the local-rated railroads were re-classified as either National III or National IV, a new grade introduced in the 2012 revision. In light of these, we assign the design speed of National III in the 2012 revision to Local I-rated railroads and National IV to Local II, and III rated railroads.

Lastly, similar to the roads built before 1994, we use the rating in 1994 to infer the design speeds of the existing railroads in 1994. The logic is the same as before: if an existing railroad was designated a particular rating in 1994, its quality must be closest to that rate. Otherwise, the railroad would have received a different rating.

Definition of Terrains As shown in Table 1, the design speed of the roads and railroads differ by four types of terrain: the plains (Ping Yuan), the low rolling hills (LRH, Wei Qiu), the hills (Zhong Qiu), and the mountains (Shan Ling). The four types of terrains are defined in the *Land Regulations in Highway Engineering*, published by the Ministry of Transportation. Panel (a) of Figure B.3 summarizes the definition of terrains used in this paper, and Panel (b) maps the terrains in China as defined. In this part, we provide more details about the definition of terrains.

Chapter 3.0.3 in the *Land Regulations* defines the terrains. Both plains and LRH are the areas with a local elevation range (LER) of less than 200 meters; Plains are further defined as the areas where the slopes are smaller than 3 degrees, and the low rolling hills as the areas where the slopes are between 3 and 20 degrees.

The definition of “hills” and “mountains” are less clear. the *Land Regulations* group the “hills” and the “mountains” into one category (Zhong Qiu Shan Ling) and defines them as the terrains with greater than 20 degrees of slope, or with LER greater than 200 meters. This aggregation of hills and mountains is because, by the time of the publication of the *Land Regulations* in 1999, the Ministry of Transportation no longer distinguishes between these

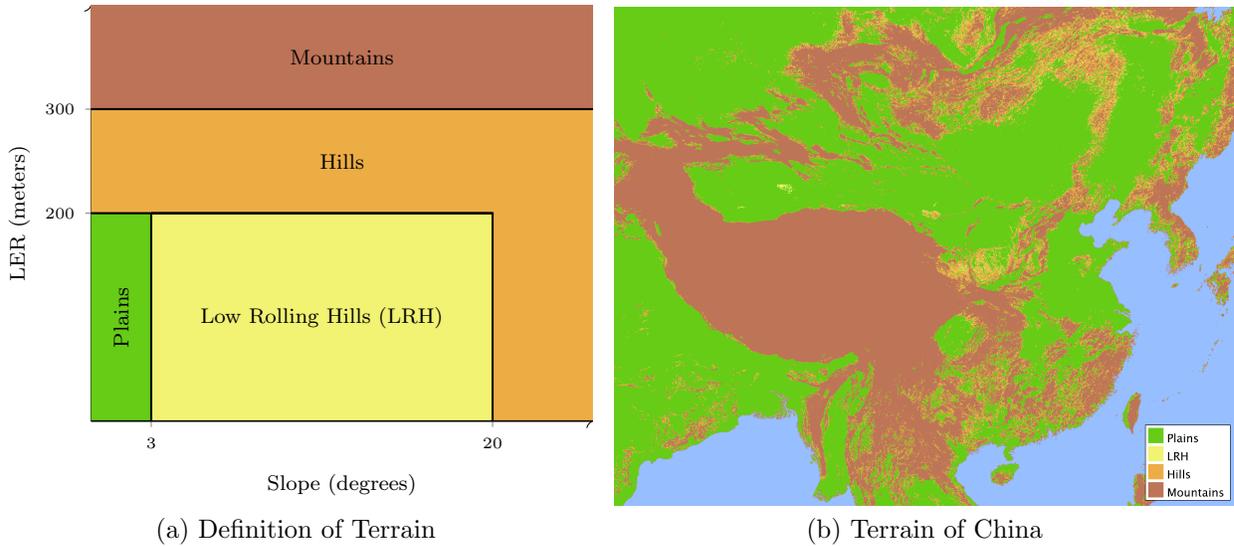


Figure B.3: Definition of Terrains, and the Terrain of China

Notes: Panel (a) defines the four types of terrains by slope and local elevation range (LER), conditional on elevation smaller than or equal to 2500 meters. All the areas with an elevation greater than 2500 meters are defined as “mountains”. Panel (b) shows the terrain of China following this definition.

two types of terrains. The Ministry also publishes the official definition of the technical terms in the *Standard of Technical Terms for Highway Engineering*. Unfortunately, the distinction between “hills” and “mountains” are subjective and non-quantitative in this official publication. The “hills” are defined as the areas with relatively large LER but without characteristics of a mountain, such as a ridge, peak, or base. The “mountains”, on the other hand, are the terrains with large LER and observable characteristics. To quantify the design speed by terrain, we need a systematic way to distinguish between the hills and the mountains. The need to distinguish these two terrains arises because, in the 1988 revision of the standards, the design speed of highways differs between the hills and the mountains, as seen in Table 1.

To separately define hills and mountains, we turn to the official definition used by the UN Environment Programme World Conservation Monitoring Centre (UNEP-WCMC), which defines “mountains” as those with a LER greater than 300 meters or with an elevation greater than 2500 meters, following [Kapos et al. \(2000\)](#). Taking the *Land Regulations* and the UNEP definitions together, we define the “hills” as the terrains with more than 20 degrees of slope or with an LER between 200 and 300 meters, and the “mountains” as those with

more than 300 meters in LER, or more than 2500 meters in elevation.

We use the data from GTOPO30 from the USGS for the slope and the LER data. The LER is then inferred from the elevation data. The slope and LER were evaluated for a five-pixel radius following [Kapos et al. \(2000\)](#).

B.3 Data Sources:

In addition to the maps and publications mentioned above, we have also utilized the following datasets:

1. The **GTOPO30** from USGS. We use this data source to define the terrains as specified above. In particular, we use four sets of data to cover China: E060N40, E060N90, E100N40, and E100N90. This dataset provides the slope and elevation data with a grid spacing of 30 arc seconds, which is approximately 1 kilometer.
2. The **One Percent Population Survey** in 2005 and 2015. We use these two datasets to estimate the aggregate stay rates between 2000 and 2005, and between 2010 and 2015. In both years, we define the aggregate stay rate as the fraction of individuals whose current location is the same as the self-reported location five years ago.
3. The **Custom Transaction Dataset** between 2000 and 2005. This dataset is used to identify the coastal prefectures that directly trade with the international markets.
4. The **City Statistical Yearbooks**. This dataset provides the data on the goods and the passenger transportation through each prefecture by mode of transportation. We drop air traffic in passenger transportation as it is only responsible for a small fraction of the total traffic.
5. The **Statistical Yearbooks** of various provinces. We used these data to construct the initial population and output in 1994. Many prefectures in our 291-prefecture sample were upgraded from “counties” after 1994 and therefore were not covered by the City Statistical Yearbook back then. We reconstruct the population and output of these prefectures using the county-level data from the provincial-level Statistical Yearbooks.

6. The **Investment Climate Survey** conducted by the World Bank in 2002. We use the firm-level information to estimate the fraction of out-of-prefecture sales of Chinese firms, the same as in [Ma and Tang \(2020\)](#).
7. The **AIS Vessel Data**. This dataset reports the location, speed, and heading of vessels. We extract the speed of moving vessels, excluding fixed platforms, military vessels, and policy vessels. We use these data to approximate the average sailing speed.
8. The **Hukou Index** are constructed using Peking University Law Information Database (<http://www.lawinfochina.com/>). Following [Fan \(2019\)](#), we collect the laws and regulations implemented at prefecture-level that are potentially related to hukou reform by searching a set of keywords including any combination of “hukou” or “huji” (also means hukou) with “gaige” (reform) or “guanli” (management), together with “chengshihua” or “chengzhenhua” (both mean urbanization) and “luohu” or “ruhu” (both mean granting hukou). The same scoring system (on a scale of 0-6 by adding up the subscores obtained from central districts and other parts of a city) and criteria (depending on housing tenure status and the length of contributing to local social security) are also applied as in [Fan \(2019\)](#) to our 291 prefectures.

C Details of the Model

C.1 Dynamic Discrete Choice

Recall that the dynamic discrete choice problem defined in the main text, replicated here for reference:

$$v_{jt} = \log \left(\phi_{jt} \frac{w_{jt}}{P_{jt}} \right) + \max_{i, r^m \in \cup_{m=1}^M \mathcal{R}_{ij}^m} \left\{ \delta V_{i,t+1} - \bar{d}_{it} - \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}} \right) + \kappa \cdot \varepsilon_{it, r^m} \right\}, \quad (\text{C.1})$$

Different from [Caliendo et al. \(2019\)](#), the individuals choose a destination i and a route to go from j to i among M types of transportation modes, $r^m \in \cup_{m=1}^M \mathcal{R}_{ij}^m$.

Denote $\zeta_{it, r^m} = \delta V_{i,t+1} - \bar{d}_{it} - \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}} \right) + \kappa \cdot \varepsilon_{it, r^m}$. Recall that ε_{it, r^m} follows a Generalized Type-I Extreme Value Distribution (GEV-I) with location parameter $\bar{\gamma}$ and a

scale parameter of 1. As ζ_{it,r^m} is a linear transformation of ε_{it,r^m} , it is also a GEV-I with a location parameter $\left[\delta V_{i,t+1} - \bar{d}_{it} - \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}}\right)\right] + \kappa\bar{\gamma}$ and a scale parameter κ , because GEV-I is closed under linear transformations.

In light of this, the maximization problem in Equation (C.2) is equivalent to maximizing over countably many ζ_{it,r^m} with a common scale parameter κ and different location parameters:

$$v_{jt} = \log\left(\phi_{jt} \frac{w_{jt}}{P_{jt}}\right) + \max_{i,r^m \in \cup_{m=1}^M \mathcal{R}_{ij}^m} \{\zeta_{it,r^m}\}. \quad (\text{C.2})$$

As GEV-I is also closed under maximization, it is straightforward to see that the distribution function of “ $\max_{i,r^m \in \cup_{m=1}^M \mathcal{R}_{ij}^m} \{\zeta_{it,r^m}\}$ ” must be:

$$\begin{aligned} \Pr\left(\max_{i,r^m \in \cup_{m=1}^M \mathcal{R}_{ij}^m} \{\zeta_{it,r^m}\} < x\right) &= \prod_{i=1}^J \prod_{r^m \in \cup_{m=1}^M \mathcal{R}_{ij}^m} \Pr(\zeta_{it,r^m} < x) \\ &= \prod_{i=1}^J \prod_{m=1}^M \prod_{r^m \in \mathcal{R}_{ij}^m} \exp\left(-\exp\left(-\frac{x - \left[\delta V_{i,t+1} - \bar{d}_{it} - \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}}\right) - \kappa\bar{\gamma}\right]}{\kappa}\right)\right) \\ &= \exp\left(-\sum_{i=1}^J \sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \exp\left(-\frac{x - \left[\delta V_{i,t+1} - \bar{d}_{it} - \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}}\right) - \kappa\bar{\gamma}\right]}{\kappa}\right)\right) \end{aligned}$$

Notice that:

$$\begin{aligned} &\sum_{i=1}^J \sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \exp\left(-\frac{x - \left[\delta V_{i,t+1} - \bar{d}_{it} - \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}}\right) - \kappa\bar{\gamma}\right]}{\kappa}\right) \\ &= \exp\left\{\log\left[\sum_{i=1}^J \sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \exp\left(-\left(\frac{x - \kappa\bar{\gamma}}{\kappa} - \frac{\left[\delta V_{i,t+1} - \bar{d}_{it} - \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}}\right)\right]}{\kappa}\right)\right)\right]\right\} \\ &= \exp\left\{\log\left[\exp\left(-\frac{x - \kappa\bar{\gamma}}{\kappa}\right) \sum_{i=1}^J \sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \exp\left(\frac{\left[\delta V_{i,t+1} - \bar{d}_{it} - \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}}\right)\right]}{\kappa}\right)\right]\right\} \\ &= \exp\left\{-\frac{x - \kappa\bar{\gamma}}{\kappa} + \log\left[\sum_{i=1}^J \sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \exp\left(\frac{\left[\delta V_{i,t+1} - \bar{d}_{it} - \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}}\right)\right]}{\kappa}\right)\right]\right\} \\ &= \exp\left\{-\frac{x - \kappa\bar{\gamma} - \kappa \log\left[\sum_{i=1}^J \sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \exp\left(\frac{\left[\delta V_{i,t+1} - \bar{d}_{it} - \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}}\right)\right]}{\kappa}\right)\right]}{\kappa}\right\}. \end{aligned}$$

Therefore, the distribution of $\max_{i,r^m \in \cup_{m=1}^M \mathcal{R}_{ij}^m} \{\zeta_{it,r^m}\}$ is another GEV-I with the location parameter:

$$\kappa \log \left[\sum_{i=1}^J \sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \exp \left(\frac{\delta V_{i,t+1} - \bar{d}_{it} - \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}} \right)}{\kappa} \right) \right] + \kappa \bar{\gamma},$$

and the scale parameter κ . It directly follows from the property of GEV-I that:

$$\begin{aligned} \mathbb{E} \left[\max_{i,r^m \in \cup_{m=1}^M \mathcal{R}_{ij}^m} \{\zeta_{it,r^m}\} \right] &= \kappa \log \left[\sum_{i=1}^J \sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \exp \left(\frac{\delta V_{i,t+1} - \bar{d}_{it} - \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}} \right)}{\kappa} \right) \right] \\ &= \kappa \log \left[\sum_{i=1}^J \exp \left(\frac{\delta V_{i,t+1}}{\kappa} \right) \sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \exp \left(-\frac{\bar{d}_{it} + \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}} \right)}{\kappa} \right) \right]. \end{aligned}$$

Now, define λ_{ijt} as:

$$\begin{aligned} \lambda_{ijt} &= -\kappa \log \left[\sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \exp \left(-\frac{\bar{d}_{it} + \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}} \right)}{\kappa} \right) \right] \\ &= -\kappa \log \left[\exp \left(-\frac{\bar{d}_{it}}{\kappa} \right) \sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \exp \left(-\frac{\left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}} \right)}{\kappa} \right) \right] \\ &= \bar{d}_{it} - \kappa \log \left[\sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \exp \left(-\frac{\left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}} \right)}{\kappa} \right) \right], \end{aligned}$$

which is the expected migration costs across all the possible routes and modes that start from j , conditional on moving to location i , the same as in the main text in equation (8).

We can then re-write the expectation term as:

$$\begin{aligned} \mathbb{E} \left[\max_{i,r^m \in \cup_{m=1}^M \mathcal{R}_{ij}^m} \{\zeta_{it,r^m}\} \right] &= \kappa \log \left[\sum_{i=1}^J \exp \left(\frac{\delta V_{i,t+1}}{\kappa} \right) \exp \left(-\frac{\lambda_{ijt}}{\kappa} \right) \right] \\ &= \kappa \log \left[\sum_{i=1}^J \exp \left(\frac{\delta V_{i,t+1} - \lambda_{ijt}}{\kappa} \right) \right], \end{aligned}$$

and therefore, we arrive at equation (7) in the main text:

$$\mathbb{E}(v_{jt}) = V_{jt} = \log \left(\phi_{jt} \frac{w_{jt}}{P_{jt}} \right) + \kappa \log \left[\sum_{i=1}^J \exp \left(\frac{\delta V_{i,t+1} - \lambda_{ijt}}{\kappa} \right) \right].$$

C.2 Solution of λ_{ijt}

Starting from the definition of λ_{ijt} as in equation (8), we follow the steps in the Appendix D.1 of [Allen and Arkolakis \(2022\)](#) and extend their methods to allow for multiple modes of transportation. In particular, we enumerate the set \mathcal{R}_{ij}^m by the length of routes, denoted by K . Denote the set of all routes from j to i under mode m with length K as \mathcal{R}_{ij}^{mK} , we first re-write λ_{ijt} as:

$$\begin{aligned} \lambda_{ijt} &= \bar{d}_{it} - \kappa \log \left[\sum_{m=1}^M \sum_{r^m \in \mathcal{R}_{ij}^m} \exp \left(- \frac{\left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}} \right)}{\kappa} \right) \right] \\ &= \bar{d}_{it} - \kappa \log \left[\sum_{m=1}^M \sum_{K=0}^{\infty} \sum_{r^m \in \mathcal{R}_{ij}^{mK}} \exp \left(- \frac{1}{\kappa} \sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}} \right) \right] \\ &= \bar{d}_{it} - \kappa \log \left\{ \sum_{m=1}^M \sum_{K=0}^{\infty} \sum_{r^m \in \mathcal{R}_{ij}^{mK}} \left[\prod_{k=1}^K \exp \left(- \frac{d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}}}{\kappa} \right) \right] \right\}. \end{aligned}$$

Note that the last step converted the additive transportation costs in $d_{(\cdot)}^{m\mathbb{P}}/\kappa$ into multiplicative costs in the unit of $\exp \left(\frac{d_{(\cdot)}^{m\mathbb{P}}}{\kappa} \right)$, similar to the main text in [Allen and Arkolakis \(2022\)](#). To proceed, we denote the adjacency matrix as $\mathbf{F}_t^{m\mathbb{P}}$, where $F_{ijt}^{m\mathbb{P}} = \exp \left(- \frac{d_{ijt}^{m\mathbb{P}}}{\kappa} \right)$ is the (i, j) th element of the matrix, and $(\mathbf{F}_t^{m\mathbb{P}})^K$ is the matrix raised to the power K . Note that we can enumerate all the routes with length K :

$$\begin{aligned} \lambda_{ijt} &= \bar{d}_{it} - \kappa \log \left\{ \sum_{m=1}^M \sum_{K=0}^{\infty} \left[\sum_{k_1=1}^J \sum_{k_2=1}^J \cdots \sum_{k_{K-1}=1}^J \left(F_{i,k_1,t}^{m\mathbb{P}} \times F_{i,k_2,t}^{m\mathbb{P}} \cdots \times F_{k_{K-1},j,t}^{m\mathbb{P}} \right) \right] \right\} \\ &= \bar{d}_{it} - \kappa \log \left\{ \sum_{m=1}^M \sum_{K=0}^{\infty} (\mathbf{F}_t^{m\mathbb{P}})_{ij}^K \right\}, \end{aligned}$$

where $(\mathbf{F}_t^{m\mathbb{P}})^K$ is the (i, j) th element of the adjacency matrix raised to the power K . Define $\mathbf{B}_t^{m\mathbb{P}}$ as the Leontief Inverse of $\mathbf{F}_t^{m\mathbb{P}}$:

$$\mathbf{B}_t^{m\mathbb{P}} \equiv (\mathbf{I} - \mathbf{F}_t^{m\mathbb{P}})^{-1} = \sum_{K=0}^{\infty} (\mathbf{F}_t^{m\mathbb{P}})^K.$$

We then arrive at the solution of λ_{ijt} as shown in the main text:

$$\lambda_{ijt} = \bar{d}_{it} - \kappa \log \left(\sum_{m=1}^M b_{ijt}^{m\mathbb{P}} \right),$$

where $b_{ijt}^{m\mathbb{P}}$ is the (i, j) th element of the matrix $\mathbf{B}_t^{m\mathbb{P}}$.

C.3 Solving the Model in Levels

Conditional on observing $\{L_{j0}\}$, Θ and a finite sequence of $\Theta_t, t = 1, 2, \dots, T$, we solve the model in levels as follows.

1. Assume that the fundamentals are constants after period T , so that $\Theta_{T+k} = \Theta_t, \forall k > 0$.
2. Solve the stationary equilibrium after T , denoted as $\bar{\Upsilon} = \{w_j, L_j, P_j, V_j, \mu\}$ from the following system of equations:

$$w_j L_j = \sum_{i=0}^J \frac{(w_j \tau_{ij})^{-\theta} (\bar{A}_j L_j^\alpha)^\theta}{\sum_{k=1}^J (w_k \tau_{ik})^{-\theta} (A_k L_k^\alpha)^\theta} w_i L_i \quad (\text{C.3})$$

$$P_j = \Gamma (1 + (1 - \eta/\theta)) \left(\sum_{i=1}^J (w_i \tau_{ji})^{-\theta} (A_i L_i^\alpha)^\theta \right)^{-1/\theta} \quad (\text{C.4})$$

$$V_j = \log \left(\bar{\phi}_j L_j^\beta \frac{w_j}{P_j} \right) + \kappa \log \left(\sum_{i=1}^J \exp(\delta V_i - \lambda_{ij})^{1/\kappa} \right) \quad (\text{C.5})$$

$$\mu_{ij} = \frac{\exp(\delta V_i - \lambda_{ij})^{1/\kappa}}{\sum_{i'=1}^J \exp(\delta V_{i'} - \lambda_{i'j})^{1/\kappa}} \quad (\text{C.6})$$

$$L_i = \sum_{j=1}^J \mu_{ij} L_j \quad (\text{C.7})$$

To solve this system:

- (a) Guess $\{L_j\}$
- (b) Solve $\{w_j\}$ from equation (C.3).
- (c) Compute $\{P_j\}$ using $\{w_j, L_j\}$ and equation (C.4).
- (d) Solve $\{V_j\}$ from equation (C.5). Note that in order to efficiently compute V_j , re-write equation (C.5) as:

$$\begin{aligned}
V_j &= \log \left(\bar{\phi}_j L_j^\beta \frac{w_j}{P_j} \right) + \kappa \log \left(\sum_{i=1}^J \exp \left(\frac{\delta V_i}{\kappa} - \frac{\lambda_{ij}}{\kappa} \right) \right) \\
&= \log \left(\bar{\phi}_j L_j^\beta \frac{w_j}{P_j} \right) + \kappa \log \left(\sum_{i=1}^J \exp \left(\frac{\delta V_i}{\kappa} \right) \times \exp \left(-\frac{\lambda_{ij}}{\kappa} \right) \right)
\end{aligned}$$

In this way, we can evaluate the sum inside the second log function as the inner product between two vectors, $\{\exp(\frac{\delta V_i}{\kappa})\}_{i=1}^J$ and $\{\exp(-\frac{\lambda_{ij}}{\kappa})\}$. Importantly, the second vector is not a function of V_i , and therefore can be evaluated outside of the loop.

- (e) Compute μ from equation (C.6).
 - (f) Update $\{L_j\}$ from equation (C.7). Repeat until converge.
3. Pick a $T' > T$ and guess a sequence of $\{V_{jt}\}_{t=1}^{T'}$, where $V_{jT'} = V_j$ as computed from above.
 4. With the $\{V_{jt}\}_{t=1}^{T'}$ and the initial population, we can compute a sequence of $\{\mu_{jt}\}_{t=1}^{T'}$ and $\{L_{jt}\}_{t=1}^{T'}$ using equations (9) and (10) forward from $t = 1, \dots, T'$. In particular, re-write equation (9) as:

$$\mu_{ijt} = \frac{\exp \left(\frac{\delta V_{i,t+1}}{\kappa} \right) \exp \left(-\frac{\lambda_{ijt}}{\kappa} \right)}{\sum_{i'=1}^J \exp \left(\frac{\delta V_{i',t+1}}{\kappa} \right) \exp \left(-\frac{\lambda_{i'jt}}{\kappa} \right)}$$

5. With the sequence of $\{L_{jt}\}_{t=1}^{T'}$, solve the sequence of $\{w_{jt}\}_{t=1}^{T'}$ using equation (18).
6. Compute the sequence of $\{P_{jt}\}_{t=1}^{T'}$ using equation (19)

7. Update the sequence of $\{V_{jt}\}_{t=1}^{T'}$ backwards from $T' - 1$ to $t = 1$ using equation (7):

$$V_{jt} = \log \left(\phi_{jt} \frac{w_{jt}}{P_{jt}} \right) + \kappa \log \left(\sum_{i=1}^J \exp \left(\frac{\delta V_{i,t+1}}{\kappa} \right) \exp \left(-\frac{\lambda_{ijt}}{\kappa} \right) \right).$$

8. Repeat until convergence.

C.4 Moment Conditions

Trade Shares by Mode The volume of sales from j to i via mode m is:

$$\sum_{r^m \in \mathcal{R}_{ij}^m} \pi_{ijt,r^m} X_{ij} = \frac{(w_{jt}/A_{jt})^{-\theta} \sum_{K=1}^{\infty} (\mathbf{F}_t^{m\text{f}})^K}{\sum_{k=0}^J (w_{kt}/A_{kt})^{-\theta} \tau_{ikt}^{-\theta}} X_{ij} = \frac{(w_{jt}/A_{jt})^{-\theta} b_{ijt}^{m\text{f}}}{\sum_{k=0}^J (w_{kt}/A_{kt})^{-\theta} \tau_{ikt}^{-\theta}} X_{ij},$$

The share of sales from j to i via mode m can be written as:

$$s_{ijt}^{m\text{f}} = \frac{\frac{(w_{jt}/A_{jt})^{-\theta} b_{ijt}^m X_{ij}}{\sum_{k=0}^J (w_{kt}/A_{kt})^{-\theta} \tau_{ikt}^{-\theta}}}{\sum_{m'=1}^M \frac{(w_{jt}/A_{jt})^{-\theta} b_{ijt}^{m'} X_{ij}}{\sum_{k=0}^J (w_{kt}/A_{kt})^{-\theta} \tau_{ikt}^{-\theta}}} = \frac{b_{ijt}^{m\text{f}}}{\sum_{m'=1}^M b_{ijt}^{m'\text{f}}},$$

and the share of all the sales from j via mode m , denoted as $s_{jt}^{m\text{f}}$ is the weighted average of $s_{ijt}^{m\text{f}}$ across destinations excluding ROW, where the weight is X_{ijt}/X_{jt} :

$$\begin{aligned} s_{jt}^{m\text{f}} &= \sum_{i=1}^J s_{ijt}^{m\text{f}} \frac{X_{ijt}}{X_{jt}} \\ &= \sum_{i=1}^J \left(\frac{b_{ijt}^{m\text{f}}}{\sum_{m'=1}^M b_{ijt}^{m'\text{f}}} \right) \frac{X_{ijt}}{X_{jt}}. \end{aligned}$$

Migration Shares by Mode The fraction of migrants from j to i via mode m at time t is:

$$\begin{aligned}
s_{ijt}^{m\mathbb{P}} &= \frac{\sum_{r^m \in \mathcal{R}_{ij}^m} \exp\left(\frac{\delta V_{i,t+1} - \bar{d}_{it} - \left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}}\right)}{\kappa}\right)}{\sum_{m'=1}^M \sum_{r^{m'} \in \mathcal{R}_{ij}^{m'}} \exp\left(\frac{\delta V_{i,t+1} - \bar{d}_{it} - \left(\sum_{k=1}^K d_{r_{k-1}^{m'}, r_k^{m'}, t}^{m'\mathbb{P}}\right)}{\kappa}\right)} \\
&= \frac{\sum_{r^m \in \mathcal{R}_{ij}^m} \exp\left(\frac{-\left(\sum_{k=1}^K d_{r_{k-1}^m, r_k^m, t}^{m\mathbb{P}}\right)}{\kappa}\right)}{\sum_{m'=1}^M \sum_{r^{m'} \in \mathcal{R}_{ij}^{m'}} \exp\left(\frac{-\left(\sum_{k=1}^K d_{r_{k-1}^{m'}, r_k^{m'}, t}^{m'\mathbb{P}}\right)}{\kappa}\right)} \\
&= \frac{\sum_{K=0}^{\infty} (\mathbf{F}_t^{m\mathbb{P}})_{ij}^K}{\sum_{m'=1}^M (\mathbf{F}_t^{m'\mathbb{P}})_{ij}^K} \\
&= \frac{b_{ijt}^{m\mathbb{P}}}{\sum_{m'=1}^M b_{ijt}^{m'\mathbb{P}}}.
\end{aligned}$$

Therefore the total out-migration from j to all the other (non-ROW) locations via mode m is the weighted average of the above probability, where the weight is the migration flow:

$$s_{jt}^{m\mathbb{P}} = \sum_{i=1}^J \left(\frac{b_{ijt}^{m\mathbb{P}}}{\sum_{m'=1}^M b_{ijt}^{m'\mathbb{P}}} \right) \mu_{ijt} = \sum_{i=1}^J \left(\frac{b_{ijt}^{m\mathbb{P}}}{\sum_{m'=1}^M b_{ijt}^{m'\mathbb{P}}} \right) \frac{L_{ijt}}{L_{jt}}.$$