The Impact of Transportation Networks on Internal Migration Flows in China

by

Su (Sam) Wang

Submitted to the Distinguished Majors Program

Department of Economics

University of Virginia

April 30, 2021

Advisor: John E. McLaren

The Impact of Transportation Networks on Internal Migration Flows in China

Su (Sam) Wang*

University of Virginia

April 2021

Abstract

This paper uses a simple gravity-type location choice model to study the effect of transportation networks on the internal migrant flows in China. After calculating the "effective distance" of inter-provincial migration using the detailed highway system map of China, I estimate that the "effective distance" elasticity of migration is around -1.45. The result implies that holding other factors constant, 1% change in such measurement will lead to around 1.45% increase in the share of migrant flows. Simple calculations show that the improvement in road networks should increase out-migration rates by about 27%. This study adds to the literature that studies internal migration frictions in China by further breaking down the cost of internal migration in China. It also quantifies the impact of transportation infrastructure investments on labor mobility within countries.

* Helpful comments from Prof. John McLaren, Prof. Amalia Miller, and fellow DMP students are gratefully acknowledged. I would also like to thank my friends Xiangmo Chen and Xinyu Hu for assisting me with the search for the data. The responsibility for remaining errors is mine.

1 Introduction

Internal migration, especially rural-to-urban migration, is an extremely prominent and key economic phenomenon in China in recent decades. The internal migration flows in China are different from their counterparts around the world because of their massive scale and predominantly cyclic nature. It is estimated that about 250-300 million people migrated every year in the last decade.[1] These workers are commonly referred to as *liudong renkou*, or "floating population." As the name implies, their movement is not permanent. Most of them come from the towns and villages in the western inland provinces, move to cities on the eastern and southeastern shore of China, and then return home at the end of the year. These internal migrants are crucial to achieving an efficient spatial equilibrium of labor inside an economy. Studies have found that migration flows, along with trade liberalization, promote the growth in aggregate productivity (Poncet 2006; Tombe and Zhu 2019).

Analyzing internal migration within China is unique also because of the system of household registration called *hukou*, which has been in place since the 1950s. In this system, everyone in a household is registered at a specific location and is entitled only to certain public resources, including pre-tertiary education and healthcare (government welfare programs), at their registration location. This introduced significant frictions to labor mobility and led to the extremely low rate of rural-urban migration in China compared to the rest of the world before the "opening-up" in the 1980s.

Besides the significant frictions brought by the *hukou* system, transportation infrastructure is also an important part of the migration costs. The origins of most of the internal migrants, the western inland provinces, have rugged topography and less-developed transportation networks. Road and railway connections to the neighboring cities are crucial for people who wish to migrate. Building these infrastructures not only facilitates the movement of people but also promotes the spread of information. People coming in and out can bring useful information about economic opportunities in other cities, and thus induce even more

[1] According to *China Statistical Yearbook* 2019.

people to migrate, whereas people in insulated regions are less likely to move because they know nothing about the prospects of living in another place.

This DMP thesis attempts to break down the specific factors of migration costs and quantify the effect of transportation networks in China, specifically the national expressway network, on the overall migrant flows. The goal of the project is to estimate the distance elasticity of migration, where the distance is defined as the "effective distance" between cities when taking into account the existing transportation networks. It adds to the literature that studies migration determinants and frictions by further investigating factors that affect the cost of internal migration in China. This area of research, although rich with literature, seldom looks at the development of transportation networks over time and studies the time-series variation of infrastructures. This project can also help us understand the economic impact of transportation infrastructure investments, other than facilitating domestic trade (Coşar and Demir 2016).

1.1 Internal Migration in China

With the restrictions imposed by the *hukou* system and related policies, labor mobility in China was largely constrained before the 1980s. After the country reformed its economy and started to "open up," the numbers of internal migrants in China started to increase rapidly. Since 2000, the central and local governments also started to deepening their reform of the preexisting *hukou* system, further easing policy restrictions to allow for more movement across cities. From the start of the century to 2015, the total number of migrant workers—those who stay and work for a sufficiently long time in a year outside of their household registration location—has more than doubled, increasing from 144 million in 2000 to 294 million in 2015.[2]

These migration flows are sometimes called "transient populations" because of their cyclic nature. Quite commonly, people move into the cities at the start of the lunar year and go

[2]See Chan (2013), Chan and Yang (2020) for more detailed summaries of the recent trends in internal migration.

back to their hometowns at the end of that year. This is partly due to cultural reasons: the spring festival, just like Thanksgiving and Christmas, is the most important holiday in China during which people return to their hometowns and enjoy time with family and relatives. One could also argue that the household registration system still in place leads to the annual cycle. Although people are allowed to move freely across cities, the policy frictions on healthcare, education, and sometimes the housing market access prevent people from re-settling permanently in cities with their entire family.

1.2 Transportation Infrastructure in China

Along with easing policy restrictions, China underwent significant development in its transportation infrastructure in the same period. Before the year 1984, China's mainland had no modern expressways. The most advanced roads were national highways, or *guodao*, which are not limited-access roads, followed by provincial roads and other local roads. The first expressway project to begin construction in mainland China is the Shenyang–Dalian Expressway in 1984, whereas the first one to be completed is the Shanghai–Jiading Expressway in 1988. Four more years later, in 1992, the Ministry of Transportation announced the plan to construct an expressway network, officially known as the National Trunk Highway System (Sloboda and Yao 2007). At the start of the century, the mainland provinces of China had about 16,300 kilometers (10,200 miles) of expressways in total, with an average density of 2.3708 kilometers per thousand km^2 of used land.[3]

Later, in January 2005, the Ministry of Transportation introduced the "7-9-18" expressway planning, envisioning a grid of 34 expressways spanning China that includes 7 radial expressways out from Beijing, 9 north-south expressways (down), and 18 east-west expressways (across). Two other north-south expressways, the Hohhot-Behai expressway and the

[3]Here, used land is defined as lands that are either used in agriculture (cultivated) or in "construction"—for building, manufacturing, mining, transportation networks, and irrigation systems. The exact definition was pinned down by the Ministry of Land and Resources in 2001 and was later expanded to a more detailed classification in 2007. The figures for 2000 were not reported in the 2001 *China Statistical Yearbook*, so I use the 2003 numbers from the 2004 yearbook, which is the closest year with such figures reported. I intend to use this area measure to approximate the area of habitable lands instead of the total land area.

Yinchuan-Baise expressway, were later added to the planned network in 2013, making the final plan a "7-11-18" network.[4] By the year 2015, total expressway mileage increased to 123,500 kilometers (77,200 miles), showing a more than sixfold increase in 15 years, or a 14.45% increase per year if compounded annually. The national average density also increased to 18.0576 kilometers per thousand km^2 of used land.

On the other hand, the growth of railway length throughout this period was less drastic compared to that of the road network, due to the large amount of fixed capital required in railroad construction. In 2000, the mainland provinces of China had about 68,700 kilometers (42,900 miles) of railroads in use, averaging 8.5240 kilometers per thousand km^2 of used land. Within 15 years, railroad length increased by 76.13% to 121,000 kilometers (75,630 miles), arriving at an average intensity of 17.6844 kilometers per thousand km^2 of used land.

Most of the railroad improvements in this period were on the transport capacity. From 1997 to 2007, the speeds of trains were raised in six waves, from a national average of 48.1 kilometers per hour to 70.18 kilometers per hour. Previous single-track railways were expanded into double tracks. Along the railroad main lines, passenger and freight trains get separated to increase the overall efficiency. China also pushed for large-scale railroad electrification, raising the electrification rate from 21% in 1995 to 75% in 2016.[5] These improvements led to significant reductions in fuel consumption and increases in railroad efficiency.

The most important development in the Chinese railway industry in this period is the introduction of High-speed Rail (HSR), which are rail lines with a designed speed of 200–350 km/h, about two to three times faster than the usual "express trains." In 2008, the Beijing-Tianjin intercity rail became the first HSR line opened in China. By 2015, the total length of HSR lines exceeded 20,000 kilometers. This figure rose to 37,900 km by the end of 2020.[6]

[4]Source (in Chinese): Expressways of People's Republic of China. http://www.owlapps.net/owlapps_apps/articles?id=91854&lang=zh.

[5]Source: Global Sustainable Electricity Partnership. "Electrification of transportation infrastructure in China improves efficiency." https://globalelectricity.org/case-studies/electrification-of-transportation-infrastructure-in-china-improves-efficiency.

[6]See Lawrence, Bullock, and Liu (2019) for more detailed information on China's HSR networks.

1.3 Previous Works

This paper relates to numerous works on the effect and implications of internal migration in China. Early on, scholars looked at the various determinants of migration decisions, including age, education, gender, marriage, and land allocation.[7] A recent work by Hao, Sun, Tombe, and Zhu (2020) studied the effect of internal migration in China from the year 2000 to 2015 on aggregate productivity. They modeled the migration costs as a migration utility discount factor and a loss of land rebates (share of the land fixed factor income among local residents of the same *hukou* registration), and they found at least 18% of the increase in aggregate GDP in China can be attributed to migration cost reductions. Other scholars in the field also studied the effect of internal migration flows on inter-generational human capital accumulation and local fiscal policies (Sieg, Yoon, and Zhang 2020), fertility decisions (Liao, Wang, Wang, and Yip 2020), and urban wages of workers with different education levels (Combes, Démurger, Li, and Wang 2020).

This paper also relates to the literature on the economic impact of transportation infrastructure improvements. Scholars in the field of international and inter-regional trade found strong positive relationships between transportation network improvements and trade volumes (for example, Coşar and Demir 2014).[8] On the other hand, in the field of labor, Heuermann and Schmieder (2019) studied the two waves of German High-Speed Rail construction and found that "a reduction in travel time by 1% raises the number of commuters between regions by 0.25%." Poot, Alimi, Cameron, and Maré (2016) looked at changing "distances" between locations in New Zealand due to transportation infrastructure improvement and new speed laws, but they did not find a statistically significant effect on migration flows. Monte, Redding, and Rossi-Hansberg (2018) developed a general equilibrium model and estimated that a reduction of commuting costs across the country generated welfare gains at

[7]See Zhao (2004)'s summary of the field studying internal migration in China. For a summary of the broader field of internal migration, see Etzo (2008) and Lagakos (2020).

[8]Also, see Redding (2020) for an excellent summary of recent researches on geography and trade. This survey paper also includes sections on transport infrastructure and measurement of geographic trade costs.

around 3.3%. Morten and Oliveira (2016) studied the effects of roads on trade and migration in the case of Brasilia, a planned city by the government, also using a general equilibrium model. They found that the road improvement increased welfare by 13.3%, of which 95% was due to reduced trade costs and 5% to reduced migration costs.

In the context of China, Poncet (2006) estimated the effect of wages, unemployment, and distance-related costs of migration on rural-urban migration between 1985 and 1995 at the provincial level. She found similar distance elasticities of migration over time and significant border effects. However, she used only simple road distances between provincial capitals without considering the different transportation quality and designs. Banerjee, Duflo, and Qian (2012) studied the effect of transportation network access on GDP per capita and its growth, but, in their simple model, they assumed labor to be immobile. Baum-Snow et al.(2017) examined roads and railways and their significant effects on the decentralization of urban economic activities. Since their focus is within cities, they used counts of radial roads, ring roads, and highway mileage to measure transportation networks. Baum-Snow et al.(2020) found highway constructions slow the growth of hinterland prefectures, but their research focused on cross-sectional estimates instead of the time-series variation of transportation networks.

Other researchers also looked at the economic impact of High-Speed Rail connections, commonly employing a difference-in-differences framework and finding induced economic migration towards larger and higher-tier cities (Xu and Sun 2020) and decreased wage of local rural-urban migrants (Kong, Liu, and Yang 2019). Another work by Ma and Tang (2020a) used a structural general equilibrium model to analyze the local welfare effect of internal migration from 2000 to 2005, while considering the effect of transportation networks on endogenous population movement. The authors digitized a high-resolution map with transportation networks following the method established by Allen and Arkolakis (2014). Their later working paper (Ma and Tang 2020b) built on this framework and constructed a panel of transportation networks from 1994-2017. They found that the transportation network induced migration and increased regional inequality. However, they seemed to use only 2000-2005 migration flow to calibrate the parameters and used 1995-2016 city population growth rates to estimate and simulate the flows.

The social and economic impacts of the migration flows and transportation infrastructure improvements are well observed by the rich set of literature. However, less work has been done in the field studying the interaction between transportation networks and migration flows in China. This project adds to these aforementioned works by looking at the effect of transportation networks on inducing internal migration in China and attempts to estimate the distance elasticity of migration weighted by the transportation network.

In the next section, I will describe the data sources and provide initial data summaries. Section 3 delineates the empirical model and estimation strategy. Section 4 presents the key results of the paper, and section 5 concludes.

2 Data

2.1 Migration Flows

In this paper, I look at inter-provincial migration, as defined by people having a different residence as their *hukou* registration province. The relevant numbers of these bilateral migration flows are reported in the 2000 and 2010 decennial censuses, as well as the 2005 and 2015 1% population sample, which cover the mainland 31 provinces (or provincial administrative regions). The sample statistics of the out-migration rate in each province are summarized in Table 1, all weighted by the corresponding sample size. The average migration rate increased from 29% in 2000 to around 33% in later years. Apart from the year 2005, when the standard deviation of the out-migration rate spiked, the regional dispersion of out-migration rates seemed to decrease over time. The distribution moved from right-skewed to more symmetric and centered, evident from the closing gap between the medians and means, as well as between the first quartile and the third quartile.

Table 1: Descriptive Statistics of Out-Migration Rates

Sample Year	Sample Total	Mean	Standard deviation	Q1	Median	Q3
2000	144,390,748	0.2938	0.1790	0.1465	0.2577	0.4653
2005	1,945,881	0.3401	0.1978	0.1797	0.3283	0.5326
2010	260,937,942	0.3291	0.1653	0.1978	0.3367	0.4847
2015	4,545,854	0.3315	0.1425	0.2126	0.3458	0.4795

Throughout the years, Guangdong province had the lowest inter-provincial out-migration rates around 4%, and only increased to 10.1% in 2015. Anhui province had the highest out-migration rates in all samples: 56 to 60% of people with *hukou* in Anhui live in a different province. Figure 1 shows the average out-migration rates for each province over the four survey periods. People from the middle and western provinces tend to be inter-provincial migrants, whereas people from eastern coastal provinces are less likely to migrate to other provinces.

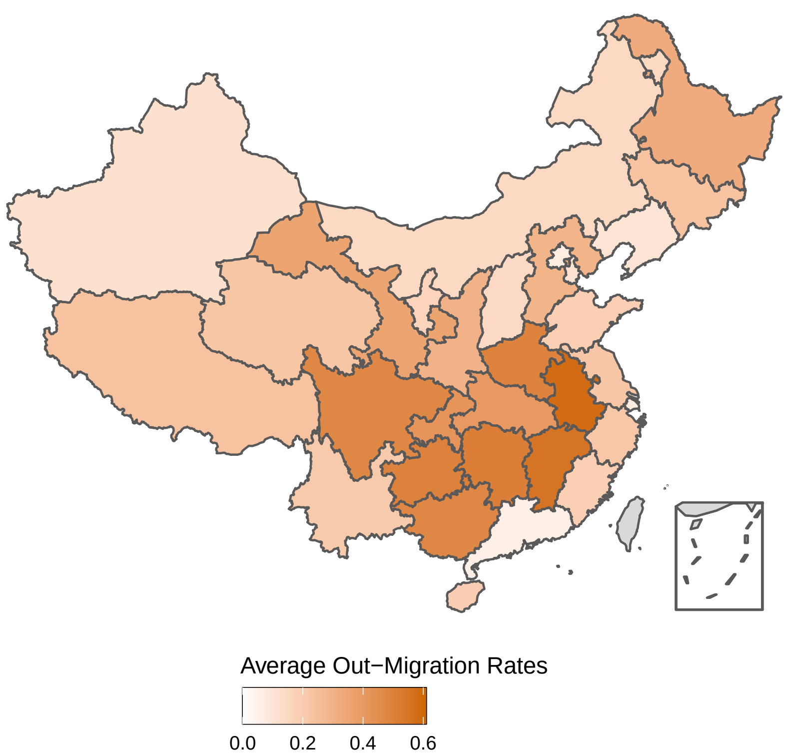

Figure 1: Average Out-Migration Rates for each Province

2.2 Transportation Networks

In this paper, I focus on the road networks due to their significance in passenger transport (Ma and Tang 2020a): the total number of passengers traveling by road is usually several to more than a hundred times than by rail in each city.[9] The GIS data on the Chinese road networks, especially for older years, are difficult to obtain, and as a result, I can look at the highway networks only in the year 2005 and 2015. More recent network data can be found on Open Street Map (OSM), which is the source of 2015 road network data for this project.[10] The 2005 road data is modified from the road shapefiles for the *China 2000 & 2010 County Population Census Data with GIS Maps* provided by the China Data Center. The original files, especially for expressways, exhibited several temporal inconsistencies.[11] Therefore, I correct the shapefiles by cross-referencing the completion dates of highways, keeping only sections that were open to traffic by January 2005.

Table 2: Unit Travel Costs for Different Types of Roads

Classification	**Common Speed Limit**	**Unit Cost**
Expressway	120 km/h (75 mph)	1
National Road (*Guodao*)	80 km/h (50 mph)	1.5
Provincial Road (*Shengdao*)	60 km/h (37.5 mph)	2
Other local road	50 km/h (31.25 mph)	2.4

After cleaning the shapefiles, I assign a unit travel cost to each road in the network based on its classification and speed limit. Table 2 shows the relationship between road classification, usual speed limit, and assigned unit cost. The cost is an inverse of the speed limit and hence should be directly proportional to the time required to travel along the road. Ma and Tang (2020a) also used similar strategies to assign unit costs for highways,

[9]Source: *China City Statistical Yearbook*.

[10]The mainland China road network extract, made by Geofabrik, is downloadable at the following link: http://download.geofabrik.de/asia/china.html

[11]For example, the files showed a few sections of expressways in Hunan province that were open to traffic after 2007, but they did not show certain sections of expressways that were already open to use by 2005.

but they used only three levels of road classifications. I obtain the "effective distances" used in empirical estimation by accumulating these unit costs along the least-cost paths between the reference points.[12]

This cost assignment scheme is of course only a rough approximation of the time costs of traveling along the roads, because the speed limits change based on the landscapes and the initial designs of the roads — highways have lower speed limits in rougher terrains, and, over time, the standards of highways improve significantly, making later ones much easier to traverse than earlier ones. However, it is difficult to check for these details using the current dataset.

To calculate the least-cost distances between provinces, I need a reference location (the departure and destination point) for each province. The most obvious cities to choose are the capitals of the provinces, which are often important cities in their province. People usually move to these cities from other areas within the province and migrate to other provinces (usually to capital cities as well) from there.

However, capitals are not always the most important point of departure and destination, so I also use two alternative sets of reference points to calculate the distances between provinces: one group is the set of the most populous city in each province, and another is the set of cities with most passengers traveling on roads. These statistics can be found in the *China City Statistical Yearbook*'s.[13] For the first measure, I use the central city population instead of the total population in the prefecture, because inter-provincial migrants usually move to the urban center rather than to the outlying counties and villages. This criterion gives me capital cities as reference points for all provinces except two — Baotou, Inner Mongolia gets chosen instead of Hohhot, and Putian, Fujian instead of Fuzhou. As

[12] For simplicity, I use the default shape length in ArcGIS to measure road lengths. Therefore, the unit cost here should be interpreted as the cost to travel through a shape with one unit of length in the software under the WGS 1984 coordinate system. Shape length times the corresponding unit cost gives me the total cost of traversing the entire road. Half of the total cost is added if the least-cost path goes through only half of the road.

[13] For the years 2006 and 2016, respectively. In cases where there is a disagreement between the Yearbooks, the city with a higher average is selected.

expected, this distance measure is highly correlated to the "capital distances" above (correlation coefficient being 0.9993 for 2005 distances and 0.9994 for 2015 distances). The second set of cities differs more from the capitals, but it still resulted in a highly correlated distance measure (0.9725 for 2005 distances and 0.9798 for 2015 distances).[14] I will refer to these two distance measures as the "Population" distance measure and the "Road Passenger" distance measure, respectively, in the tables hereafter.

I refrain from using reference points such as population and migrant population centers of gravity of the province because of the way I calculate the "effective distances" using the transportation network. Two large cities wide apart within a province may lead to a population center of gravity in the middle of the province, where the transportation network is less developed. In that case, the shortest path to some other province might have to go through the two big cities, and the resulting "average" distance could be larger than the distances calculated from the two large cities.

Table 3: Descriptive Statistics of Average Effective Distances

Sample Year	Measure 1: Capitals		Measure 2: Population		Measure 3: Road Passenger	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
2005	22.1405	7.8133	22.2252	7.7893	23.2934	7.7530
2015	18.2900	6.3924	18.3424	6.3741	19.1105	6.3250

Table 3 summarizes some relevant statistics for all three distance measures. For each province and distance measure, I average the effective distances to all other provinces and report the means and standard deviations of the resulting 31 numbers, for the years 2005 and 2015 separately. Over time, we see reductions in the mean and variability of average distances in all three measures. The average reduction in distances between 2005 and 2015

[14]The full list of cities for this distance measure is as follows: Beijing; Tianjin; Shijiazhuang, Hebei; Yuncheng, Shanxi; Baotou, Inner Mongolia; Shenyang, Liaoning; Changchun, Jilin; Harbin, Heilongjiang; Shanghai; Suzhou, Jiangsu; Wenzhou, Zhejiang; Lu'an, Anhui; Fuzhou, Fujian; Jiujiang, Jiangxi; Yantai, Shandong; Zhumadian, Henan; Wuhan, Hubei; Hengyang, Hunan; Guangzhou, Guangdong; Wuzhou, Guangxi; Haikou, Hainan; Chongqing; Chengdu, Sichuan; Zunyi, Guizhou; Kunming, Yunnan; Lhasa, Tibet; Xi'an, Shaanxi; Jiuquan, Gansu; Xining, Qinghai; Yinchuan, Ningxia; and Urumqi, Xinjiang.

is around 18%, obtained by regressing the 2015 distances on 2005 distances and suppressing the constant.

Figure 2 looks at the reduction in "capital distances" specifically. Most outlying provinces experienced a greater reduction of the average distance measures over time, consistent with the initial observations of Ma and Tang (2020b). Between 2005 and 2015, Xinjiang experienced the greatest distance reduction at 28.97%, whereas Tibet has the lowest reduction at 5.06%, likely because of the difficulty in constructing expressways on its extreme topography.

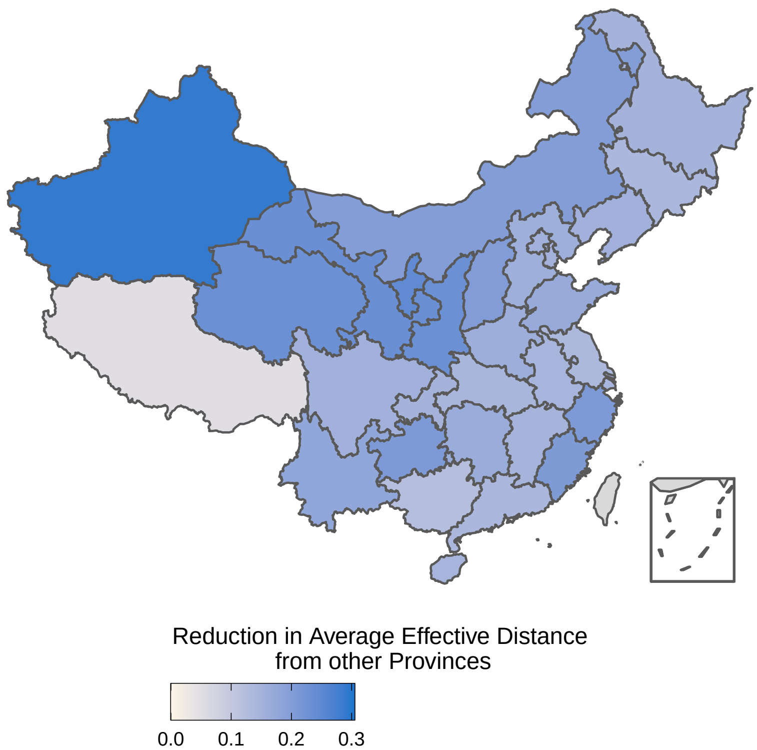

Figure 2: Reduction in Average Effective Distance away from Other Provinces (Using Capitals as Reference Points)

2.3 Province Characteristics

This study also uses wage levels and unemployment rates in each province. The relevant statistics are reported by the National Bureau of Statistics of China. The wage level is defined as the average annual income of people who have a job position (having a contract and receiving payrolls but might be temporarily on leave), and all numbers are adjusted by the CPI in the terms of 2015 Chinese Yuan.

Since the unemployment rates are missing for two provinces in 2005, I estimate the rates by dividing the number of unemployed over the total of unemployed and employed, which are found in the 2005 and 2015 1% population surveys. Table 4 summarizes the relevant statistics for wage levels and unemployment rates in each province in 2005 and 2015, weighted by destination sample size.

Table 4: Descriptive Statistics of Wage and Unemployment Rates

Sample Year	Sample Total	Wage (2015 RMB)		Unemployment Rate (%)	
		Mean	Standard deviation	Mean	Standard deviation
2005	1,945,881	24,741.46	6,953.63	3.2613	1.4839
2015	4,545,854	64,220.61	15,842.01	3.1923	0.6575

3 Empirical Framework

I use a gravity-type estimation of the bilateral migrant flows between prefecture-level cities in China, which is similar to the model used in Poncet (2006), but with extra terms to better describe the *hukou* system.

Consider an economy with N distinct locations, in which every worker k has the same characteristics. Each of them has an origin location $j \in \{1, 2, \ldots, N\}$ and will choose a destination $i \in \{1, 2, \ldots, N\}$ at each time period t to maximize the following individual expected utility:

$$\pi^k_{jit} = V^k_{jit} + \epsilon^k_{jit} = \rho_{it} \cdot \ln(w_{it}\delta_{it}r_{jit}d^{-\lambda}_{jit}) + (1 - \rho_{it}) \cdot u_0 + \epsilon^k_{jit}, \tag{1}$$

where ρ_{it} is the probability of finding a job at i at time t, and w_{it} the wage level at the destination location i at time t. Since I assumed that the workers are identical, these values should be the same for locals and migrants. The second term inside the logarithm, δ_{it}, encapsulates other relevant characteristics of the destination that are hard to measure, such as local amenities. The next term inside the logarithm, r_{jit}, is the term measuring the differential treatment towards migrant workers in the local labor market. I intend to use it to model policy restrictions and also anything related to common "migrant disadvantage," such as local attitudes towards migrants. I assume that this utility discount factor is the same for every migrant at i and time t, so it should only take on two values for any given destination i and time t:

$$r_{jit} = \begin{cases} 1 \,, & \text{if } j = i \\ \\ c_{it} \,, & \text{if } j \neq i \end{cases} .$$

I would expect c_{it} to be less than 1 in most cases, especially for popular destinations of internal migrants. The last term d_{jit} inside the logarithm in equation (1) is the "effective distance" between the origin j and destination i at time t, to be computed from the transportation network, as defined above in section 2.2. After the logarithm, the next term in the equation is the expected utility if the workers stay unemployed, where the Bernoulli utility function is assumed to be a constant, u_0. Finally, the last term ϵ_{jit} in equation (1) is a random "idiosyncratic utility shock" for each worker and is assumed to be from a common distribution. This term accounts for possible heterogeneous location preferences of potential migrants and thus allows for divergence in destination choices, a common approach in the literature modeling labor mobility choices (for example, Poncet 2006; Artuç and McLaren 2015).

Given the cyclic nature of the internal migration flows in China, every agent in the model would return to their origin j at the start of the next period and make their destination choice again. In this context, this means that the origin j is the household registration location, assumed to be fixed for each worker.[15]

[15] This is an extreme assumption, as it is still possible to change registration location. However, the total number of this type of "successful" migrations is too small compared to the large numbers of the floating population. These are also much more commonly used by workers with higher educational levels, who are not the majority of the internal migrants studied in this project. Therefore, this migration method should

We can expand the log in equation (1) and get

$$\pi_{jit}^k = \rho_{it}[\ln(w_{it}) + \ln(\delta_{it}) + \ln(r_{jit}) - \lambda \ln(d_{jit})] + (1 - \rho_{it}) \cdot u_0 + \epsilon_{jit}^k.$$

Again, each agent from j at time t is choosing the destination i with the highest utility π_{jik}. Therefore, if the idiosyncratic utility shock follows the extreme value distribution with the cumulative distribution function $F(\epsilon) = \exp(-\exp(-\frac{\epsilon}{\nu}))$, then the ratio of migrants originally from location j and moved to a given city i can be expressed in the following logistic form:

$$\frac{m_{jit}}{L_{jt}} = \frac{m_{jit}}{\sum_i m_{jit}} = \frac{e^{\frac{V_{jit}^k}{\nu}}}{\sum_i e^{\frac{V_{jit}^k}{\nu}}} \quad ,$$

where m_{jit} is the total number of migrants originated from j who migrated to i at time t, and L_{jt} is the total number of workers originated from j at that time. Taking logarithm of both sides of the equation, we have

$$\ln \frac{m_{jit}}{L_{jt}} = \frac{\rho_{it}}{\nu}\Big(\ln(w_{it}) + \ln(\delta_{it}) + \ln(r_{jit}) - \lambda \ln(d_{jit})\Big) + \frac{(1 - \rho_{it})u_0}{\nu} - \ln \left(\sum_i e^{\frac{V_{jit}^k}{\nu}} \right).$$

Notice that the last term of the right-hand side is summed over all possible locations i, so it should depend only on the origin city j and the time t. The second term $\ln(\delta_{it})$ depends only on the destination location i and the time period t, whereas the third term is jointly determined by the time period t, the destination location i and whether $j = i$.[16] Therefore, we can estimate the following equation using OLS:

$$\ln \frac{m_{jit}}{L_{jt}} = \rho_{it}\Big[\beta_1 \ln(w_{it}) + \beta_2 \ln(d_{jit}) + a_{it} + \ln(c_{it}) \times D_{move}\Big] + b_{jt} + \gamma_t + u_{jit}, \qquad (2)$$

where a_i, b_{jt}, and γ_t are destination, origin, and time fixed effects, D_{move} is the dummy variable taking value 1 whenever $j \neq i$, and u_{jit} the stochastic error term. Since I assume

not matter much for our purpose.

[16]Note that this term is always 0 when $j = i$.

that the coefficients β_1 and β_2 are common over time, they should measure the "average effect" of these migration factors on the overall migration distribution. The linear term for ρ_{it} coming from the unemployment part of the expected utility gets absorbed into the destination fixed effects a_{it}, whereas the constant term from it, u_0/ν, is absorbed by the time fixed effects γ_t. It is impossible to separate them in the estimation. The regression is also weighted by origin sample size to account for potential heteroskedasticity issues and to reflect the greater importance of provinces with a larger population.

I also estimate the following equation using Poisson Pseudo-Maximum Likelihood (PPML) estimation:

$$\frac{m_{jit}}{L_{jt}} = \exp\left(\rho_{it}\Big[\beta_1 \ln(w_{it}) + \beta_2 \ln(d_{jit}) + a_{it} + \ln(c_{it}) \times D_{move}\Big] + b_{jt} + \gamma_t\right) + \nu_{jit}, \qquad (3)$$

where the variables follow the same definition in equation (2), and ν_{jit} is the stochastic error term. Because of the small sample size from the 2005 and 2015 1% population survey, there are a few origin-destination pairs in the data between which there are no migrants — the logarithm is not defined for these zero-observations. The PPML approach preserves these observations and interprets these zero values as potential sampling error, introduced by the ν_{jit} error term.[17]

In the estimation of equations (2) and (3), the left-hand side comes from the bilateral migration flows. The probability of finding a job in the destination province, ρ_{it}, will be estimated by one minus the unemployment rates, and the wage term w_{it} is estimated by nominal wage levels as defined in section 2.3.

The most important variable, d_{jit}, comes from the "effective distances" described in section 2.2. However, I cannot use the logarithm of "effective distances" directly, because by definition the distance between one province and itself is 0, undefined for logarithmic function. Instead, I use an inverse hyperbolic sine transformation:

[17] See Motta (2019), Santos Silva and Tenreyro (2006) for the econometric details of PPML estimation and gravity equations.

$$\widetilde{d_{jit}} = \text{arcsinh}(d_{jit}) = \ln\left(d_{jit} + \sqrt{d_{jit}^2 + 1}\right),$$

which allows me to preserve the 0 values. $\widetilde{d_{jit}}$ is used in the estimation of equations (2) and (3), in place of $\ln(d_{jit})$. According to Bellemare and Wichman (2020), the transformation has desirable properties and negligible error when d_{jit} is large enough. This is the case in our data: if we only consider origin-destination pairs that are different (off-diagonal terms in the distance matrix), then more than 83% of them have distances greater than 10 in 2005, and still, 76% of them are so in 2015. The estimated "semi-elasticity" should be fairly close to the actual elasticity of migration.[18]

The coefficient of interest here is β_2, the effective distance elasticity of migration flows. It tells us that with 1% smaller in "effective distances", whether geographically due to shorter distance or temporally due to improved transportation, the percentage increase in migrant flows between the two provinces, holding other characteristics constant. Under the assumptions of the functional form of the utility π and the structure of the idiosyncratic preference shock ϵ, the expected value for $\widehat{\beta}_1$ should be $1/\nu$ and the expected value of $\widehat{\beta}_2$ should be $-\lambda/\nu$. Therefore, it is also possible to back out λ, the distance exponent in the utility function, from the estimates, by dividing $-\widehat{\beta}_2$ over $\widehat{\beta}_1$.

4 Results

4.1 Distance Elasticity of Migration

The main regression results are summarized in Table 5. The estimates for the distance elasticity are fairly consistent across distance measures and estimation methods, being significantly negative, usually between -1.52 and -1.43. The estimates from PPML estimation tend to be slightly smaller in magnitude than the corresponding OLS estimates, and the

[18]The actual elasticity should be $\beta_2 \cdot \frac{d}{\sqrt{d^2+1}}$, which is close to β_2 when d is large. The semi-elasticity estimate deviates from the true elasticity more when d is smaller. When $d = 0$, the elasticity is 0.

Table 5

Wage and Distance Elasticities of Migration

	(1)	(2)	(3)	(4)	(5)	(6)
$\rho\times$Wage	0.3310	4.5250***	0.2779	4.4775***	-0.2697	3.2496***
	(2.8426)	(1.1496)	(2.8579)	(1.1497)	(2.9681)	(0.9707)
$\rho\times$Distance 1	−1.5161***	−1.4855***				
(Capital)	(0.0398)	(0.0490)				
$\rho\times$Distance 2			−1.5075***	−1.4818***		
(Population)			(0.0401)	(0.0488)		
$\rho\times$Distance 3					−1.4601***	−1.4346***
(Road Passenger)					(0.0424)	(0.0441)
Estimation Method						
OLS	✓		✓		✓	
PPML		✓		✓		✓
Summary Statistics						
N	1,892	1,922	1,892	1,922	1,892	1,922
R^2	0.8656	-	0.8642	-	0.8535	-
Pseudo-R^2	-	0.5168	-	0.5167	-	0.5168

[i] Robust standard errors are reported for PPML estimates.

third distance measure yields the most conservative results.[19] In general, holding all other factors constant and assuming that people can always find a job, 1% reduction in the effective distance increases the share of migrants between the given origin and destination provinces by about 1.45%. This estimate is slightly larger in magnitude compared to the road distance elasticity around -1 in Poncet (2006), suggesting greater effects of distances on migrant flows in later years.

This coefficient estimate can be interpreted in two ways: spatial or temporal. However, due to the modeling assumption, the estimate measures only the distance-induced effect on migration, and hence, we should interpret the practical impacts of road networks with

[19] As seen in equations (2) and (3), the coefficients are not for the variables (wage or distance) alone, but the product of ρ and the corresponding variable. Therefore, the elasticity also depends on the probability of finding a job in the destination. This result comes from the modeling assumption: in the model, the Bernoulli utility of unemployment does not depend on the "effective distance" measure. The overall elasticity, which is a combination of the two elasticities of Bernoulli utility terms, thus becomes smaller in magnitude when more weight is put on the unemployment term.

caution. If we take the spatial perspective, for a given origin, the ratio of people moving to a nearer destination is affected by not only the distances between the origin and destination, but also the destination fixed effects a_{it} and destination-specific migration barriers c_{it}. From the temporal perspective, for a given origin and destination pair, the time fixed effects also contribute to the difference in migration shares between 2005 and 2015, not just the reduction in effective distances. Therefore, in both ways, the distance-induced effects on migration are only part of the story.

With that in mind, we can still do a simple back-of-the-envelope calculation of the increase in out-migration rates when nothing else changes. Regressing the 2015 distances on 2005 distances and suppressing the intercept term, we can get the average reduction rate for each effective distance measure. Multiplying these rates with the corresponding elasticity estimate, we get that a uniform reduction in distances at the average level, without any other changes, will increase out-migration rates between 2005 and 2015 by 27.17%, 27.17%, and 27.01%, respectively.[20] The estimates suggest that the improvement in road networks is more than adequate to explain the observed rise in mobility in this period. Given the partial equilibrium approach of this paper, it is impossible to go deeper and decompose precisely the actual changes of worker distributions inside the economy.[21]

As shown in columns (1), (3), and (5), the linear log-log specification gives insignificant and at times negative estimates of the wage elasticity of migration. PPML estimation, on the other hand, gives significantly positive estimates of the coefficient, which allows us to back out λ with greater precision.[22] Table 6 shows the point estimates and 95% confidence intervals of

[20]The calculation cannot be extended to the stayers (origin the same as destination) because the inverse hyperbolic sine transformation leads to greater estimation error when the distance is small. The actual distance elasticity for those "staying" observations is always 0 — see footnote 18. This characteristic prevents the following paradox from happening: the shares of people migrating from j to i increases for all destination i following the reduction in distances, when they should add up to 1 all the time.

[21]From a general equilibrium viewpoint, increased road access leads to greater labor mobility, which also makes the wage differentials between regions smaller—this will, in turn, decrease overall migration rates. Maybe that is one reason why we do not see a dramatic rise in out-migration rates from 2005 to 2015 in the data.

[22]The wage coefficients seem to imply a variance parameter ν of the extreme value distribution between 0.14 and 0.74, centered around 0.25. This value is smaller than comparable estimates in other labor mobility studies (for example, Artuç and McLaren (2015) estimated it to be between 0.5 and 1.6.), suggesting that

λ from each distance measure, following the method outlined by Lye and Hirschberg (2018). The most conservative range of estimates for λ, given by distance measure 3, is between 0.18 and 0.70.

Table 6
Estimates of λ for Different Distance Measures

Distance Measures	Point Estimate	Std. Err.	[95% Conf. Interval]
Capital	0.3283	0.0836	[0.1645, 0.4921]
Population	0.3309	0.0852	[0.1639, 0.4980]
Road Passenger	0.4415	0.1313	[0.1842, 0.6988]

4.2 Migration Barriers

Another interesting result from the model is the estimates of the fixed effects $\ln(c_{it})$, which usually can be interpreted as migration barriers into the destination i at time t. The rate of change of the migrant utility discount rate over time, $\ln(c_{i,2015}) - \ln(c_{i,2005})$, exhibits interesting geographical patterns.

Table 7: Summary Statistics of Migration Barrier Reduction

Distance Measure	**Mean**	**Standard deviation**	**Q1**	**Median**	**Q3**
Capital	0.0310	0.5479	-0.3266	-0.0140	0.4490
Population	0.0307	0.5484	-0.3265	-0.0164	0.4435
Road Passenger	0.0393	0.5153	-0.3247	0.0406	0.4662

The distributions of the migration barrier reduction, arising from the regression of each distance measure, are summarized in table 7. They all seem to be centered closely around 0, and the first two distributions are skewed slightly to the right.

Figure 3 shows the reduction in migration barriers between 2005 and 2015, averaged over three PPML estimates. In the figure, negative values stand for increases in $\ln(c_{it})$.

heterogeneous individual preferences have a smaller dispersion in this context. People will likely respond more to wage differentials across regions, and the shocks on wages (for example, export boom in coastal regions) get transmitted throughout the economy to a wider range of regions.

The changes in migrant barriers exhibit large spatial heterogeneity. Most eastern coastal provinces experienced a large reduction in barriers, except for Shandong, Zhejiang, and Fujian province. Shanghai leads with a decrease of 1.30, indicating that without other factors changing, the reduction in barriers in Shanghai induces about 130% more migrant shares towards Shanghai from every other province. Tianjin follows with a reduction of 0.98, and Beijing with 0.89.

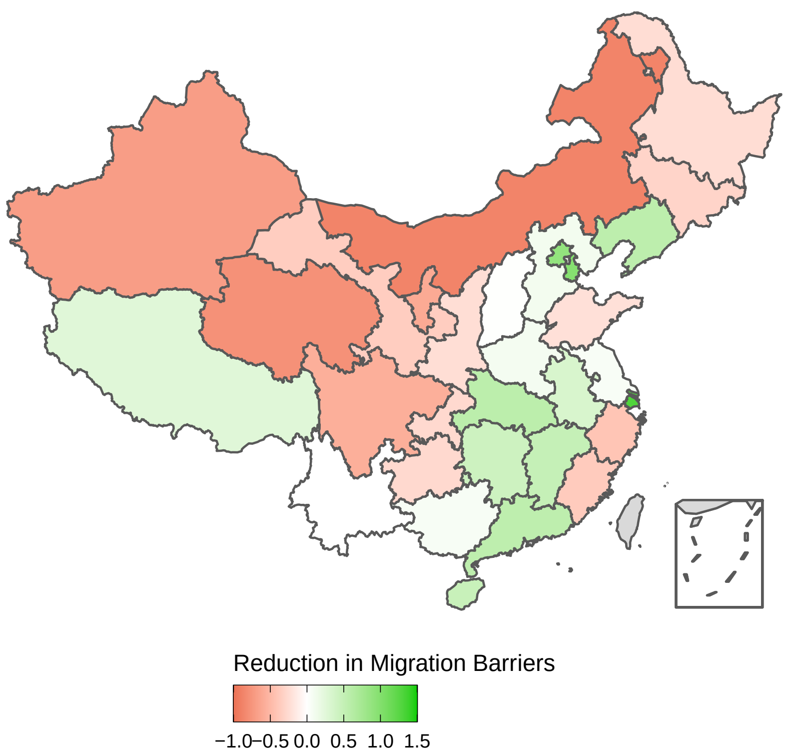

Figure 3: Average Reduction in Migration Barriers between 2005 and 2015

On the other hand, western and inland provinces have rising "barriers" during this time period. The reason is probably not that those provinces tightened their migration policies significantly, but that these destinations become less attractive to migrants. One speculation is that since the barriers in "better" destinations, such as Beijing and Shanghai, lowered significantly, people can move to those provinces with greater ease and no longer feel the need to stay at closer and merely "acceptable" destinations—the value of staying there for migrants decreased. However, in order to examine this narrative, I will need more variables to enrich the current model.

5 Conclusion

In this DMP thesis, I analyzed the impact of "effective distances," calculated from road networks, on the shares of internal migration flows in China. The main findings of this thesis support the argument that transportation infrastructure improvement facilitates the movement of people across regions. By establishing a location-choice model incorporating unemployment, wages, and migration barriers, I found a significantly negative relationship between the distances and bilateral migrant flows — smaller "effective distances" due to nearer destinations or better expressway networks do encourage people to migrate. The results are robust with regard to different reference points within each province. Three sets of reference points agree that, on average, the "effective distance" elasticity of migrant shares is around -1.45.

Using the wage elasticity estimated from equation (4), I also estimated the distance exponent in the utility function to be between 0.17 and 0.70 with 95% confidence. I also found meaningful reductions in migration barriers in eastern provinces, in particular Beijing, Tianjin, and Shanghai, between 2005 and 2015, whereas western inland provinces experienced increases in "migration barriers."

One limitation of the study is the partial equilibrium approach employed when setting up workers' maximization problem. The agents take destination wages and unemployment rates as given, when in the real world they are endogenously determined in the economy. This is a trade-off made for the tractability of this DMP project. Establishing a more complete and complex general equilibrium model, which incorporates international and regional trade as well as factor movements, might give us an even more precise picture of the effects of transportation networks on internal migration flows.

References

[1] Allen, Treb, and Costas Arkolakis (2014). “Trade and the Topography of the Spatial Economy.” *The Quarterly Journal of Economics* 129, no. 3, 1085-1140.

[2] Artuç, Erhan, and John McLaren (2015). ”Trade policy and wage inequality: A structural analysis with occupational and sectoral mobility.” *Journal of International Economics* 97, no. 2: 278-294.

[3] Banerjee, Abhijit, Esther Duflo, and Nancy Qian (2012). “On the road: Access to transportation infrastructure and economic growth in China.” NBER Working Paper no.17897 (March).

[4] Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang (2017). “Roads, railroads, and decentralization of Chinese cities.” *Review of Economics and Statistics* 99, no. 3: 435-448.

[5] Baum-Snow, Nathaniel, J. Vernon Henderson, Matthew A. Turner, Qinghua Zhang, and Loren Brandt (2020). “Does investment in national highways help or hurt hinterland city growth?” *Journal of Urban Economics* 115: 103124.

[6] Bellemare, Marc F., and Casey J. Wichman (2020). “Elasticities and the inverse hyperbolic sine transformation.” *Oxford Bulletin of Economics and Statistics* 82, no. 1: 50-61.

[7] Chan, Kam Wing (2013). ”China: internal migration.” *The encyclopedia of global human migration.*

[8] Chan, Kam Wing, and Xiaxia Yang (2020). ”Internal migration and development: A perspective from China 1.” In *Routledge Handbook of Migration and Development*, pp. 567-584. Routledge.

[9] Combes, Pierre-Philippe, Sylvie Démurger, Shi Li, and Jianguo Wang (2020). ”Unequal migration and urbanisation gains in China.” *Journal of Development Economics* 142: 102328.

[10] Coşar, A. Kerem, and Banu Demir (2016). “Domestic Road Infrastructure and International Trade: Evidence from Turkey.” *Journal of Development Economics* 118 (January): 232–44.

[11] Etzo, Ivan (2008). ”Internal migration: a review of the literature.”

[12] Grogger, Jeffrey, and Gordon H. Hanson (2011). “Income maximization and the selection and sorting of international migrants.” *Journal of Development Economics* 95 (1), 42 – 57. Symposium on Globalization and Brain Drain.

[13] Hao, Tongtong, Ruiqi Sun, Trevor Tombe, and Xiaodong Zhu (2020). “The effect of migration policy on growth, structural change, and regional inequality in China.” *Journal of Monetary Economics*, vol. 113(C), pages 112-134.

[14] Heuermann, Daniel, and Johannes Schmieder (2019). “The effect of infrastructure on worker mobility: evidence from high-speed rail expansion in Germany.” *Journal of Economic Geography* 19, issue 2, 335-372.

[15] Kong, Dongmin, Lihua Liu, and Zhiqing Yang (2020). “High-Speed Rails and Rural-urban Migrants’ Wages.” *Economic Modelling.*

[16] Lagakos, David (2020). “Urban-rural gaps in the developing world: Does internal migration offer opportunities?” *Journal of Economic perspectives* 34, no. 3: 174-92.

[17] Lawrence, Martha, Richard Bullock, and Ziming Liu (2019). “China’s High-Speed Rail Development.” International Development in Focus, Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/31801. License: CC BY 3.0 IGO.

[18] Liao, Pei-Ju, Ping Wang, Yin-Chi Wang, and Chong Kee Yip (2020). "To Stay or to Migrate? When Becker Meets Harris-Todaro." NBER Working Paper no. 27767 (September).

[19] Lye, Jenny N., and Joe Hirschberg (2018). "Confidence Intervals for Ratios: Econometric Examples with Stata." Available at SSRN 3165266.

[20] Ma, Lin, and Yang Tang (2020a). "Geography, trade, and internal migration in China." *Journal of Urban Economics*, 115, (C).

[21] Ma, Lin, and Yang Tang (2020b). "The Distributional Impacts of Transportation Networks in China." Working Paper (May).
http://lin-ma.com/uploads/3/6/0/7/36070314/mt_transportation.pdf

[22] Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg (2018). "Commuting, migration, and local employment elasticities." *American Economic Review* 108, no. 12: 3855-90.

[23] Morten, Melanie, and Jaqueline Oliveira (2016). "The effects of roads on trade and migration: Evidence from a planned capital city." NBER Working Paper no.22158 (August).

[24] Motta, Victor (2019). "Estimating Poisson pseudo-maximum-likelihood rather than log-linear model of a log-transformed dependent variable." *RAUSP Management Journal.*

[25] Ortega, Francesc and Giovanni Peri (2013). "The Effect of Income and Immigration Policies on International Migration." *Migration Studies* 1 (1), 47–74.

[26] Poncet, Sandra (2006). "Provincial migration dynamics in China: Borders, costs and economic motivations." *Regional Science and Urban Economics*, Elsevier, vol. 36(3), pages 385-398 (May).

[27] Poot, Jacques, Omoniyi Alimi, Michael P. Cameron, and David C. Maré (2016). ”The gravity model of migration: the successful comeback of an ageing superstar in regional science.” IZA Discussion Papers No. 10329.

[28] Redding, Stephen J. (2020). ”Trade and Geography.” NBER Working Paper no. 27821 (September).

[29] Santos Silva, JMC, and Silvana Tenreyro (2006). “The log of gravity.” *The Review of Economics and Statistics* 88, no. 4: 641-658.

[30] Sieg, Holger, Chamna Yoon, and Jipeng Zhang (2020). “The Impact of Migration Controls on Urban Fiscal Policies and the Intergenerational Transmission of Human Capital in China.” NBER Working Paper no. 27764 (September).

[31] Sloboda, Brian W., and Vincent W. Yao (2007). “The Development of the Chinese Transportation Infrastructure: A Case of Highway Development.” *Journal of the Washington Institute of China Studies*, Winter, Vol. 2, No. 3: 60-72.

[32] Wang, Wei-hua, Guo-li Ou, and Chen-yang Li (2006). “Research on the Development Mechanism of China’s Highway.” In 2006 International Conference on Management Science and Engineering, pp. 2092-2097.

[33] Xu, Zhicheng, and Tianshi Sun (2020): “The Siphon effects of transportation infrastructure on internal migration: evidence from China’s HSR network.” *Applied Economics Letters*, DOI: 10.1080/13504851.2020.1796913.

[34] Zhao, Zhong (2004). “Rural-urban Migration in China-What do we know and what do we need to know?” *China Economic Quarterly*, 3: 517-536.