Urbanization Policy and Economic Development: A Quantitative Analysis of China's Differential Hukou Reforms*

Wen-Tai Hsu†　　Lin Ma‡

December 2020

Abstract

The household registration system (hukou system) in China has hampered rural-urban migration by posing large migration friction. The system has been gradually relaxed in the past few decades, but the reforms have been differential in city size. We find a striking contrast in migration patterns between years 2005 and 2015; rural people tended to move more to large cities in 2005, but more to small- and medium-sized cities in 2015. We calibrate a spatial quantitative model to the world economy in both years with China divided into rural, mega-city, and other-city regions. We find that alternative urbanization policies that are not differential and that are more laissez-faire substantially improve national welfare, in magnitudes that are comparable to the welfare gains from the trade liberalization that China has put in place in the past.

Keywords: hukou system; household registration system; differential reform; urbanization policy; economic development; spatial quantitative analysis

---

*We thank Bingjing Li and Michael Zheng Song for their help with data access to the *Intercensal Population Sample Survey of One-Percent*. We are grateful to the editor and two anonymous referees for their very helpful comments and suggestions. We also thank Davin Chor, Yizhen Gu, and the participants of the 2020 Urban Economic Association Virtual Conference and the 2020 SMU Virtual Conference on Urban and Regional Economics for comments and suggestions. The Online Appendix to this paper can be found on the journal's website, as well as the authors' personal websites, `https://wthsu.weebly.com` and `https://lin-ma.com`.

†School of Economics, Singapore Management University. Email: wentaihsu@smu.edu.sg.

‡Department of Economics, National University of Singapore. Email: ecsml@nus.edu.sg.

1 Introduction

Rural-urban migration, a.k.a., urbanization, is instrumental to the economy's industrialization process in which the structural transformation from primary to secondary/tertiary industry occurs mainly around large cities (Lewis, 1954; Harris and Todaro, 1970). In contrast with most countries where there are no significant institutional migration barriers, China presents an interesting case when studying the process of rural-urban migration, as migration friction posed mainly by the household registration system, a.k.a., *hukou system*, has significantly hampered the movement of rural labor to cities.

The hukou system is mainly a legacy of the planned-economy regime before the Reform and Opening-up in 1979, and the Chinese government and the general public are naturally aware of the various problems that stem from it. The system is complex as it involves the right to work, housing purchases, health insurance, pension, land allocation in rural areas, and access to education for the migrants' children, etc. Reforms to the hukou system have therefore been slow and somewhat difficult, and thus the system has remained until now. As we detail in Section 2, important policy documents issued by the central government such as several of the Five-Year Plans of the past two decades show that the central government tends to believe that the largest cities are too large and that their growth in size should therefore be restrained. Meanwhile, the government also recognizes the benefits of rural-urban migration for rural people and the overall economy.

It is then not surprising to find that the reforms of the hukou system have been differential in city size. Government documents have shown that thoughts about these differential reforms began to take shape around 2006, and that the reforms gradually took place during 2006–2015. The difference has intensified over time, as can be seen by comparing the two Five-Year Plans for 2006–2015. Eventually, the restrictions to obtain a hukou in small- and medium-sized cities were totally abandoned around 2014–2015. However, conditions remain for migrants seeking to obtain hukous in large cities, and it is still very difficult for the largest cities. In this paper, we seek to understand the impacts of such differential reforms.

This paper first documents a striking contrast in the migration patterns between 2005 and 2015. We find that even though there was more rural-urban migration in 2015 than in 2005, the movement to large and small cities were equal in 2005, whereas the tendency shifted drastically toward small cities in 2015. This finding is consistent with the above-mentioned reforms. We then propose a story with a global perspective in which differential reforms may be harmful to the Chinese economy compared with cases in which reforms are not differential. We calibrate a quantitative spatial model to a world economy in which China is divided into three regions: rural, mega urban region (MUR), which includes all cities with a population of at least 5 million, and other urban region (OUR), which includes all of the other cities.[1] We then conduct a quantitative analysis to evaluate these differential reforms in comparison with alternative urbanization policies. Note that our model does not qualitatively build in whether or not these differential reforms are harmful; it depends on various fundamentals calibrated from data.

For our empirical and quantitative examinations, we choose to view China as the above-mentioned three regions for the following reasons. First, the MUR is most productive in China, as it contains the major agglomeration of economic activities: the Yangtze River Delta (which includes Shanghai, Hangzhou, Suzhou, and Nanjing), the Pearl River Delta (which includes Shenzhen, Guangzhou, and Dongguan), the Beijing-Tianjin agglomeration, and many other prosperous cities. Second, we opt not to go for a many-region model within China, as we seek to sharpen the contrast between large and small cities and to allow room to incorporate other countries in the world.

Our story is briefly as follows. The differential hukou reforms may be harmful for China's economic development because when resources/population are reallocated away from the most productive region, i.e., the MUR, aggregate productivity may be lowered. In various policy narratives, the government seeks to promote industrialization in the OUR. In other words, the government prefers to move firms to the OUR over moving people to the MUR. This results in the MUR's wages rising faster than in the case without such a preference. Such policy narratives tend to ignore the fact that firms need not move to China's smaller cities if they believe that labor costs are too high. They can migrate to other developing countries (henceforth ODC) such as Vietnam for lower wages or Malaysia for better access to international trade than China's inland urban areas. Whether such unintended consequences due to international trade and firm mobility are harmful for the Chinese economy remains to be determined, because when individuals are encouraged to move to the OUR, the entry of firms there also increases. Therefore, the relative productivity and

[1] See Section 2.2 for details on the definition of cities and the partition.

wages among the ODC and China's MUR and OUR are important for determining the overall effects.

Our general-equilibrium spatial quantitative model extends that in Ma and Tang (2020b), which builds on Melitz (2003) and Tombe and Zhu (2019), to include the agricultural sector, rural regions, and a monocentric city structure within the urban regions. The model allows for multiple countries, each of which consists of multiple regions. Each region is either rural or urban; rural regions engage only in agricultural production, whereas urban regions produce differentiated products which can be interpreted to include both manufactured goods and services. As in Melitz (2003), the differentiated sector is monopolistic competitive; firms are heterogeneous in their productivity and face selection pressure. Individuals are given their initial locations, and they can choose whether to migrate to a different location based on their idiosyncratic locational preferences, migration costs, and consumption utility derived from location-specific amenity, wages, and prices of goods. A free entry (and exit) condition holds in all locations; in this sense, firms are also mobile across locations. Importantly, individual workers are mobile only within a country, but firms are mobile globally.

To properly account for urban costs such as housing and commuting costs, each urban region is modeled as a (representative) monocentric city, in which city residents choose their locations of residence and commute to the central business district (CBD) to work. In equilibrium, the longer the commute, the cheaper the housing price; overall urban costs are positively related to the city size and outside land value, which in the Chinese context can be taken as the stringency of urban land supply controlled by the city government (Du and Peiser, 2014; Hsu et al., 2017). In sum, our model allows four potential forces that could explain the observed migration pattern: migration costs, relative productivity, relative amenity, and relative conditions of the land/housing market.

Our quantitative model is calibrated to the global economy with China being divided into the three above-mentioned regions. The ODC and the rest of the world (henceforth ROW) are each divided into an urban and a rural region. Our calibration exercise involves calibrating a rich set of region-specific fundamentals by various data sources; a particularly important one is the migration-flow data, which helps pin down both migration costs and local amenity. The migration-flow data is from the 2005 and 2015 *Intercensal Population Sample Survey of One-Percent* (henceforth *One-Percent Population Survey* for short) in which surveyed individuals were asked their hukou registration and their whereabouts five years ago. Such information allows us to estimate the migration probability between each pair of Chinese regions. The initial spatial distributions of population are then obtained from the Population Census in years 2000 and 2010 for the calibration exercises in years 2005 and 2015, respectively.

The calibrated migration costs for rural-MUR and rural-OUR migration are similar in 2005; nonetheless, while both migration costs drop in 2015, the drop for rural-OUR is much sharper than that for rural-MUR, rationalizing the above-mentioned migration patterns. For the other three potential forces for explaining the observed pattern — relative productivity, relative amenity, and housing conditions — the first two are ruled out by the calibration results. If the productivity in the OUR grows faster than that in the MUR during the 2005–2015 period, i.e., if there is a regional convergence in productivity, then relative productivity may also contribute to the observed pattern. However, our calibrated results show that the relative productivity between the two regions barely changes over this 10-year span. Similarly, if amenity in the OUR relative to that in the MUR increases in this 10-year span, then this relative amenity may also contribute to the observed pattern. But we find the opposite: the relative amenity in the OUR decreases. As the calibrated housing fundamentals do reflect that the MUR's urban costs rise faster than the OUR's, we conduct a counter-factual of letting these fundamentals grow in the pace during these 10-year span. It turns out this explains only a relatively minor fraction of the observed migration pattern, compared with the counter-factual analysis based on migration costs.

Based on the baseline calibrated model in 2015, the two main counter-factual exercises that we conduct are as follows. First, we set the friction of both rural-MUR and rural-OUR migration to be the same at the level where total rural-urban emigration flows are the same as in the baseline. We find that the migration pattern is reversed compared with the observed pattern in 2015, with rural-MUR migration being larger than rural-OUR migration. This is natural because the MUR is more productive and migration costs are made the same. The resulting reallocation of labor improves national welfare by 2.6%; this magnitude is substantial considering that this involves only 4.4% of the total population reallocated from smaller cities to mega cities.

In the second counter-factual exercise, we consider a more liberalized version of the urbanization policy in which the rural-MUR migration cost is set the same as the rural-OUR cost, which remains at the baseline level. In this case, more rural people emigrate compared with the baseline, and this emigration is more toward the MUR compared with the first counter-factual. The resulting improvement in national welfare is larger, at 15.6%.

To put these welfare gains of alternative urbanization policies in perspective, we compare these gains with welfare gains from trade, as our model allows such a comparison. The question we ask is how large the percentage reductions in trade costs must be to deliver the same welfare gains from alternative urbanization policies. Corresponding to our first and second counter-factuals, we find that 5.8% and 24.6% reductions in trade costs are required. To put this in perspective, China only lowered its trade barriers by 5.1% from 1996 to 2006 according to the estimates of iceberg trade costs from ESCAP-World Bank database. This was the period when China entered the World Trade Organization (at the end of 2001) and when tariffs were substantially reduced. In other words, by adopting better urbanization policies, China could gain even more than from the trade liberalization that it has accomplished in the past.

As this paper focuses on the effect of urbanization policy on overall national welfare, it is closely related to Hsieh and Moretti (2019), who study the effect of housing constraints on discouraging labor from moving into the most productive metropolitan areas in the US. By relaxing the housing constraints in the most productive metropolitan areas and thus allowing more labor to migrate into these areas, there are substantial gains in output and welfare because of the improvement in aggregate productivity. Our paper is similar in stressing the effect of labor reallocation on aggregate productivity and welfare, but we focus on different policies, i.e., urbanization policy pertinent for developing countries. Moreover, we differ in our incorporation of international trade.

Our paper is also closely related to the literature on internal migration in China. Tombe and Zhu (2019) study the impact of migration on aggregate productivity. Fan (2019) and Zi (2020) study the effects of international trade on inter-prefecture migration and regional income inequality. Building upon Tombe and Zhu (2019), Ma and Tang (2020b) further incorporate firm entry and exit (extensive margin) and more granular internal geography than both of the above-mentioned papers. They highlight the roles of the relative locations of cities and the fact that the extensive margin could overturn the welfare results of migration counter-factuals. An et al. (2020) also focus on the recent wave of policy reforms and empirically document that the relaxation of migration restrictions in the small cities leads to higher population inflows to and lower wage rates in those cities. To the best of our knowledge, we are the first to document the contrast in patterns of rural-urban migration between 2005 and 2015. Moreover, our paper is unique as it is the first quantitative analysis on the effects of the differential hukou reforms, and we show that alternative urbanization policy can lead to substantial welfare gains, in magnitudes that are comparable to the trade liberalization that China has accomplished in the past.

Our paper is broadly related to the rapidly growing literature on spatial quantitative economics (Redding, 2016; Redding and Rossi-Hansberg, 2017) and particularly on migration and trade (Artuç et al., 2010; di Giovanni et al., 2015; Fajgelbaum et al., 2018; Caliendo et al., 2018, 2019). Lastly, our work is generally related to the literature on resource misallocation (Hsieh and Klenow, 2009; Song et al., 2011).

The rest of the paper is organized as follows. Section 2 introduces the background of the hukou system and various reforms, and presents the empirical migration patterns. Section 3 lays out the quantitative model, and Section 4 quantifies it. Section 5 conducts a quantitative analysis by considering alternative urbanization policies. Section 6 concludes.

2 Background and Motivating Facts

In this section, we first provide a background introduction of the hukou system and its evolution, highlighting the differential reforms. We then examine migration patterns in 2005 and 2015 using the *One-Percent Population Survey*.

2.1 Background

The hukou system was formally established in 1958 when the government promulgated and implemented the Ordinances on Household Registration (*Hu Kou Deng Ji Tiao Li*). Under the command-economy regime when many goods (including foods) and services were rationed, and work management planned and centralized, population control was important for the feasibility of such a regime. Therefore, the hukou system was very strict, as any migration from one location to another required various approvals and was generally disallowed.

The restrictions began to loosen after the Reform and Opening-up in 1979. In 1980s, migration to locations such as special economic zones, e.g., Shenzhen, was much easier than other locations. While the economy and labor demand in industrialized cities continued to grow, various relaxations on the hukou system have been observed during the 1990's and 2000's; they mainly vary at the level of prefectural-level cities. For examples, several prefectures allow for "blue-cover hukou" (*Lan Yin Hu Kou*) as a precursor to a regular hukou; to obtain a blue-cover hukou typically requires conditions concerning a stable job and residence, housing purchase, or investments. In some places such as Beijing and Shanghai, "work permits" (*Gong Zuo Ju Zhu Zheng*) are issued to people with certain qualifications. Oftentimes, work permits are also accompanied by point systems in which work-permit holders can obtain a regular hukou when enough points are reached. However, these precursors do not entitle the holders the same rights and benefits as regular hukous, and not every migrant is able to obtain such precursors.

Not having a hukou or having only a precursor affects one's employment opportunities, access to health insurance, unemployment insurance, children's right to public schools, and pensions even when one has a formal, full-time job. Universities set different thresholds for the college entrance exam for different hukous, and they usually favor local hukous; this makes local hukous in hubs of higher education such as Beijing and Shanghai all the more valuable.

Even though the Chinese government is aware of the need to reform the hukou system in order to reduce migration friction, their reforms are differential in city size. The evolution of the chapters regarding urbanization and regional planning in the Five-Year Plans, which are announced in the first year of every five-year period as comprehensive guidelines for what the central government plans to do and their priorities in the coming five years, sheds light on the changes in policy and the underlying thought process. In the 10th Five-Year Plan, which covers 2001–2005, Chapter 9 briefly discusses urbanization policy and mentions the "coordinated development of large, medium, and small sized cities". This chapter also discusses reforming the hukou system and suggests revoking unreasonable restrictions for rural people to migrate and work in cities and towns. But that is all: it does not mention any other details on the reform.

In the 11th Five-Year Plan, which covers 2006–2010, Chapter 21 provides more details about how the rights of rural migrants in cities and towns should be protected, and how hukous should be gradually granted to those who have stable jobs and residences. Importantly, this chapter explicitly states that rural migrants should be encouraged to move into small and medium sized cities and small towns, while the population growth in mega cities should be controlled and contained by industrial means.

In the 12th Five-Year Plan, which covers 2011–2015, Chapter 20 puts even more stress than the previous Five-Year Plan on the reform of the hukou system. It says that the sizes of mega cities are to be "controlled", population "management" of large and medium sized cities is to be improved, and small and medium sized cities and small towns are to relax their conditions for obtaining hukou. As a result of such a plan, the State Council of China released an official document in 2014 titled *Opinions on Further Reforms on the Household Registration System*, which echoes what was outlined in the 12th Five-Year Plan, but asks all towns and small cities (of which the population is below 500,000) to totally abandon restrictions on obtaining hukous. The same Opinion also mentions "strictly controlling" the size of mega cities (of which the population is above 5 million); the opinions on the large and medium-sized cities lie somewhere in between. Regional planning and, in their terminology, "optimizing the structure of city size distribution", have become so important that the government also announced a *National New Type Urbanization Plan* in 2014 that covers 2014–2020. The plan is very detailed on almost every aspect of urban planning, regional planning, and city development; again, the plan explicitly states that "the main goal of optimizing the structure of city size distribution is to speed up the development of small and medium sized cities" (Chapter 12).

Based on the concerns over regional inequality, all three Five-Year Plans discuss the Western Development Program; in the latter two Five-Year Plans, strategies to develop the Middle Region are also explicitly discussed. All of these mention the potential and strategies of industrialization in these non-coastal regions, which contain mostly small- and medium-sized cities and only a few mega cities.

2.2 Migration Patterns

Cities in this paper are defined to be the urban districts of a prefecture (*Shi Xia Qu*) which approximate metropolitan areas by international standard (Fujita et al., 2004). The population of these urban districts within a prefecture is therefore the urban/city population of the prefecture. To sharpen both our empirical examination and quantitative analysis, we divide China into three regions: rural, mega urban region (MUR), and other urban region (OUR). Since the 2014 *Opinions on Further Reforms on the Household Registration System* by the State Council dictates that the sizes of those cities with an urban population of at least 5 million are to be strictly controlled, we choose this 5-million urban population as the threshold by which to divide urban China. Thus, those 21 cities above this threshold make up the MUR, whereas all of the other cities are collectively the OUR.[2] As prefectures partition China, the rural areas in all of the prefectures are thus collectively the rural region.

We examine migration patterns in 2005 and 2015 using the information on migration flows from the *One-Percent Population Surveys* in 2005 and 2015. This population inter-censal survey is part of China's census program. The survey was an attempt to cover one percent of the total population, and utilized a stratified multi-stage cluster sampling process. Given China's total population, these are very large samples. Surveyed individuals were asked their current locations, hukou registrations, and their whereabouts five years ago. One can use either hukou registrations or the locations five years ago as the origins for migrants; the resulting migration flows are highly correlated between the two measures in terms of prefecture-to-prefecture migration. However, because the 2005 survey reveals the surveyed individuals' locations five years ago only at the prefecture level but not any finer, we are unable to determine whether their origins were rural or urban. Thus, we adopt only the definition of origin by hukou registration, which tells whether a hukou is rural or urban. For more data details, see Online Appendix C.

Let the migration probability, m_{ij}, be defined as the probability of an individual from location j moving to location i. The survey allows us to compute the *observed* migration probability $\tilde{m}_{ij}$ by dividing the numbers of individuals from location j who move to i by the total number of individuals from location j.

Table 1 shows the matrices of migration probability in 2005 and 2015. In 2005, 2.9% of rural individuals migrated to the MUR, whereas 2.8% of them migrated to the OUR. In 2015, these numbers rose sharply to 8.4% and 29.4%, respectively. The large increases in the migration probabilities from rural to both urban regions indicate easier rural-urban migration and suggest that the relaxation of the hukou restrictions does work to promote urbanization. Nevertheless, whereas

[2]Note that the MUR also corresponds to the top two tiers of cities per the official definition of the State Council (document reference number 000014349/2014-00135), which classifies prefectures into 7 tiers according to their urban population. The top tier (Chao Da) includes the prefectures with an urban population of at least 10 million, and the second tier (Te Da) includes those with an urban population of at least 5 million and below 10 million. Online Appendix Table A.1 lists these 21 cities. The Online Appendix to this paper can be found on the journal's website, as well as the authors' personal websites, `https://wthsu.weebly.com` and `https://lin-ma.com`.

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.944	0.033	0.070
MUR (d)	0.029	0.959	0.024
OUR (d)	0.028	0.008	0.907

(a) Migration Probability, 2005

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.622	0.078	0.180
MUR (d)	0.084	0.894	0.079
OUR (d)	0.294	0.027	0.742

(b) Migration Probability, 2015

Table 1: Matrices of Migration Probability

Note: This table presents the matrices of migration probability. An element at the i-th row and the j-th column indicates the probability of an individual originating from j and moving to i. Each column sums to 1. The data source is the *One-Percent Population Survey* in the respective years, and an "origin" is defined as the place of hukou registration.

the migration probability to the large cities was similar to that to the smaller cities in 2005, it became only a fraction of the probability to the OUR in 2015. Such a striking reversal suggests that the effects of the differential reforms on the hukou system were substantial, causing the rural population to increasingly move toward the smaller cities. Also evident from comparing the two matrices is that the staying probabilities are all lower in 2015 than in 2005, indicating an overall increased population movement, which is consistent with relaxed migration friction.[3]

3 Model

Our model adapts the framework in Ma and Tang (2020b), which builds on Melitz (2003) and Tombe and Zhu (2019). In particular, we add agricultural production, rural-urban migration, amenity, and urban costs (both housing and commuting) to Ma and Tang (2020b) to study patterns of rural-urban migration in China and account for the potential driving forces besides changes in migration barriers.

[3]There are similar probabilities of urban-to-rural migration to those of rural-to-urban migration in 2015. Note that this does not imply there is little net rural-urban migration. The *One-Percent Population Survey* does not offer aggregate population figures, and as the Chinese government denounced the distinction between urban and rural hukous around 2014, no public statistics reveal these population shares based on hukou after 2014. However, the urbanization rate (the fraction of actual population living in urban areas) is 54.8% (*China Statistical Yearbook*), which forms an upper bound for the urban hukou; hence the lower bound for the rural hukou is 45.2% of the total population. Simple algebra based on these numbers and Table 1 yields that the lower bound of net rural-urban migration is 5.4% of the total population in 2015. The larger the shortfall of the population share of urban hukou from the urbanization rate, the larger the net rural-urban migration flows must be.

3.1 Basic Environment

The world consists of M **countries** which we index using c or d. Each country is, in turn, made of a number of **regions** which we index as r or s. For ease of exposition, we call a country-region combination a **"location"**, and use i or j to index the locations with the understanding that i or j corresponds to a region r in country c. The total number of locations in the world is denoted as J. Each country c has population $\bar{N}_c$ which can migrate between regions within the country subject to friction specified later, but international migration is not allowed.

A region is either urban or rural. Urban regions produce differentiated products while rural regions produce a homogeneous agriculture product. We interpret differentiated products as both manufactured goods and services. Both differentiated and agricultural goods can be traded both within and across countries. We assume that intranational trade is frictionless while international trade is subject to iceberg trade costs.[4]

Another important difference between urban and rural regions is the urban costs, which primarily consist of housing costs and commuting costs. We incorporate these urban costs by embedding an (Alonso-Muth-Mills) monocentric city structure within each urban region. Even though our quantitative exercises lump multiple cities within each urban region, we will use the representative city size of each urban region to calibrate the model. Recognizing the fact that land costs (rather than construction costs) are the main driver of the differences in housing costs across different locations within a city, we opt for a monocentric-city model with land consumption only for tractability. One could, in principle, interpret the housing structure as part of the differentiated goods.

3.2 Consumption

Individuals in each location i derive utility from consumption and idiosyncratic locational preference; their utility will be discounted by migration friction if they decide to move to different locations from their original ones. The non-market components of the utility, i.e., idiosyncratic

[4]As this paper features the interaction between intranational migration and international trade, we assume frictionless intranational trade for simplicity. Also, intranational trade data are not easy to come by; as we will calibrate our international-trade aspect of the model using many countries, the lack of intranational trade data in most of these countries prevents us from incorporating intranational trade costs.

locational preferences and migration friction, will be specified later. The consumption component of the utility is given as follows.

The consumption utility function is Cobb-Douglas in the agriculture good, a CES composite of differentiated goods, and land space. For an individual living (in a representative city) in an urban region i at a distance z from the central business district (CBD), her utility function is given by

$$U_i = \frac{\alpha^{-\alpha}\gamma^{-\gamma}\left(1-\alpha-\gamma\right)^{-(1-\alpha-\gamma)}}{d\left(z\right)}\phi_i\left(y_i^A\right)^{\alpha}\left[\sum_{j=1}^{J}\int_{k\in\Omega_{ij}} y_{ij}\left(k\right)^{\frac{\varepsilon-1}{\varepsilon}}dk\right]^{\frac{\gamma\varepsilon}{\varepsilon-1}}\ell_i^{1-\alpha-\gamma}, \tag{1}$$

where ϕ_i is the amenity, $y_{ij}(k)$ is the consumption of variety k purchased from location j, $\varepsilon > 1$ represents the elasticity of substitution among varieties, y_i^A is the consumption of agriculture produce, ℓ_i is the consumption of land space, $\alpha < 1$ and $\gamma < 1$ capture the expenditure shares of agricultural and differentiated goods, respectively, $d(z) > 1$ is the "iceberg" commuting cost as in Ahlfeldt et al. (2015),[5] and Ω_{ij} denotes the set of varieties from location j available for purchase in location i and is endogenously determined.

We model the monocentric city as a disk with a radius $\bar{z}_i$ on a 2-D plane in which one can travel from any place in the disk to the CBD via a straight line. (Recall that a "location" in this paper is used to refer to a region; thus we use "place" to refer to a place within a representative city.) Production does not take land space; every individual in the city works in the CBD; the city radius $\bar{z}_i$ is endogenously determined. Individuals at different locations z face the same prices of agricultural and differentiated goods but different commuting costs $d(z)$ and land rents $R_i(z)$. In equilibrium and conditional on the idiosyncratic locational preferences, consumption utilities in different places z are equalized, which implies that the land rent $R_i(z)$ decreases in z. In textbook monocentric-city models, an outside land rent $\bar{R}_i$ is assumed to determine the city edge $\bar{z}_i$, and is often interpreted as the agricultural land rent. However, we recognize that there are institutional barriers on transfers between rural and urban land in China such as local governments' control over land supply and various stipulations on the collective ownership of rural land. Thus, $\bar{R}_i$'s are assumed to be exogenous and will be calibrated to reflect each urban region's land supply stringency; hence they will not be linked to the agricultural/rural land rent which will be endogenously

[5] Instead of the traditional approach of modeling commuting costs as monetary costs, we follow Ahlfeldt et al. (2015) by modeling the commuting costs as an iceberg cost for ease of calibration for quantitative purposes.

determined. Nevertheless, the equilibrium condition $R_i(\bar{z}_i) = \bar{R}_i$ still holds.

For the rural region, as rural residents are agriculture workers, they are assumed to live on the farms and hence there is no commuting cost, i.e., $d(z) = 1$. The land rents facing all rural residents are the same and denoted simply as R_i.

The set of varieties consumed in location i, $\Omega_i \equiv \cup_j \Omega_{ij}$, depends directly on firm entry and exit decisions in i, and also on the number of firms that choose to sell to i from all of the other locations. The entry, exit, and "exporting" decisions made by the firms are all dependent on the endogenous population distribution and migration patterns, which in turn rely on the fundamental forces in the model: sectoral productivity differences across locations, migration friction that could be affected by urbanization policies, and the trade friction that we will specify later. In general, a larger market size/access, potentially as a result of migration, supports more firms and varieties in a location, which is welfare-improving given the love of variety embedded in the utility function specification.

3.3 Differentiated Sector

We model the differentiated sector following Melitz (2003): firms with heterogeneous productivity compete in a monopolistic-competitive market, and each firm produces a unique variety. The one-to-one mapping between variety and firm allows us to interchangeably use k to index both the variety and the firm producing it.

In the differentiated sector, "exporting" from location j to i incurs a fixed cost denoted as f_{ij} in the unit of input bundles specified later. Trade is also subject to the standard iceberg trade cost denoted as $\tau_{ij} \geq 1$: to deliver one unit of a good from location j to location i, the firm must produce and ship τ_{ij} units from location j. Firms must also pay fixed costs denoted as f_{ii} units of input bundles in order to sell to the local market.

Production The production of variety k in location i is linear in the input bundles denoted as $b_i(k)$:

$$q_i(k) = \frac{1}{a(k)} b_i(k), \tag{2}$$

where $1/a(k)$ is the productivity of firm k, and input bundles are made of a Cobb-Douglas combination of local labor and a CES composite of intermediate inputs from all differentiated products available in location i:

$$b_i(k) = \beta^{-\beta}(1-\beta)^{-(1-\beta)} \left[n_i(k)\right]^{\beta} \left[\left(\sum_{j=1}^{J} \int_{k' \in \Omega_{ij}} y_{ij} \left(k'; k\right)^{\frac{\varepsilon-1}{\varepsilon}} dk' \right)^{\frac{\varepsilon}{\varepsilon-1}} \right]^{1-\beta}, \tag{3}$$

where $n_i(k)$ is the labor employment of firm k, $y_{ij}(k'; k)$ is the amount of variety k' purchased from location j for the production of k, and β is the relative weight of labor in the production function.

Firms are heterogeneous in productivity, as $a(k)$, the input bundle requirement for producing one unit of output, varies across firms. For quantitative purposes, we follow the literature and assume that firms draw productivity $(1/a)$ from a location-specific Pareto distribution:[6]

$$\text{Pr}\left(\frac{1}{a} < x\right) = 1 - \left(\frac{\mu_j}{x}\right)^{\theta},$$

where θ is the tail index and μ_j is the parameter that reflects the average productivity in location j. A higher μ_j implies that the average draw of a is lower in j, and $1/\mu_j$ defines the maximum of a. The cumulative distribution function (CDF) of a is therefore

$$G_j(a) = (\mu_j a)^{\theta}, a \in (0, 1/\mu_j].$$

Entry and Exit There is a large pool of potential entrants. To enter production in location j, an entrant must pay f_e units of input bundles acquired in location j. Upon paying the entry cost, the firm draws its productivity from $G_j(a)$, based on which it decides whether to produce or to exit. The outside option of exiting is normalized to zero.

[6]See Chaney (2008) who assumes the Pareto in a Melitz model. The Pareto distribution is often assumed because of its analytical convenience and its ability to generate power laws in firm size, which is a well-documented empirical regularity. See, for example, Luttmer (2007).

3.4 Agriculture Sector

Production in the agriculture sector requires the input bundle b_j^A, subject to a productivity parameter μ_j^A:

$$q_j^A = \mu_j^A b_j^A. \tag{4}$$

The input bundle to agriculture production is a Cobb-Douglas combination of labor (N_j), land (L_j), and intermediate goods:

$$b_j^A = \nu^{-\nu}\eta^{-\eta}\left(1-\nu-\eta\right)^{-(1-\nu-\eta)}\left(N_j\right)^{\upsilon}\left(L_j\right)^{\eta}\left[\left(\sum_{j=1}^{J}\int_{k'\in\Omega_{ij}} y_{ij}\left(k'\right)^{\frac{\varepsilon-1}{\varepsilon}}dk'\right)^{\frac{\varepsilon}{\varepsilon-1}}\right]^{1-\upsilon-\eta}. \tag{5}$$

The trade in agriculture goods is also subject to an iceberg cost denoted as $\tau_{ij}^A \geq 1$. The price of agriculture goods from rural location j in location i is then a function of productivity, trade costs, and the costs of input bundle, χ_j^A:

$$P_{ij}^A = \frac{\tau_{ij}^A \chi_j^A}{\mu_j^A}. \tag{6}$$

As the agriculture product is homogeneous, the realized price of the good in location i is the minimum of all the sellers:

$$P_i^A = \min_{j=1,\cdots,J}\left\{P_{ij}^A\right\}. \tag{7}$$

Note that the existence of fixed production factors for agricultural production[7] imply that each country's rural region must produce even if it imports from other countries. Equilibrium rural wages would adjust so that the domestic agricultural price is equal to the import price.

[7]First, the fixed land endowment must be utilized due to the Cobb-Douglas production function. Moreover, idiosyncratic locational preferences imply that there must always exist a positive labor supply in the rural region.

3.5 Rural Land Market and Land Rent Rebate

The urban land market clears in a standard way as we will illustrate in Section 3.8.2. Assume the two rural land uses (agricultural production and residential) are perfect substitutes, and hence one land rent clears the land market in each rural region. Unlike most monocentric city models which assume absentee landlords, we assume that every individual in every country c owns an equal share of the land in the country; hence the aggregate land rent in this country is rebated to each citizen evenly. The rebated amount to an individual is denoted as T_c.[8]

3.6 Migration Decision

As mentioned, there are three components in the utility of individuals: the consumption utility, and two non-market components: idiosyncratic locational preferences and bilateral migration friction. For the consumption utility, the associated indirect utility for an individual at an urban region i in country c and at a distance z from the CBD is given by

$$v_i(z) = \frac{\phi_i \left(w_i + T_c\right)}{d\left(z\right) \left(P_i^A\right)^{\alpha} \left(P_i\right)^{\gamma} R_i(z)^{1-\alpha-\gamma}}, \tag{8}$$

where w_i is the urban wage rate and P_i is the ideal price index of the differentiated goods. The standard procedure for solving the consumers' utility maximization problem yields

$$P_i = \left[\sum_{j=1}^{J} \int_{\Omega_{ij}} \left(p_{ij}(k)\right)^{1-\varepsilon} dk\right]^{\frac{1}{1-\varepsilon}}. \tag{9}$$

Furthermore, $v_i(z) = \bar{v}_i$ in equilibrium for all z within the same city. We will derive the explicit expression of $\bar{v}_i$ in Section 3.8.2. The indirect utility for an individual at a rural region i is written in the same way as that in (8) except that $d(z) = 1$ and $R_i(z)$ does not vary in z.

We now describe the two non-market components of the utility. First, each individual draws an idiosyncratic preference shock toward each location $\{\iota_i\}_{i=1}^{J}$, where ι_i is *i.i.d* across locations and

[8] Having rebates of land rents instead of assuming absentee landlords is mainly because it is better to account for all income in a general equilibrium model. For example, an increase in a particular country's urban or agricultural productivity will increase the aggregate land rents of this country, and missing these land rents will understate the income effects on various general-equilibrium objects such as relative trade flows, relative firm entry, and hence price indices among countries.

individuals. We assume that ι_i follows a Fréchet distribution with CDF

$$F(\iota_i) = \exp\left[-\left(\iota_i\right)^{-\kappa}\right],$$

where κ is the shape parameter that controls the heterogeneity of these locational-preference shocks; the smaller the κ, the larger the heterogeneity. Such locational-preference specification has been well understood as a dispersion force, which is stronger if the heterogeneity is larger (κ is smaller). See Murata (2003) for an early example.

Second, moving from j to i incurs origin-destination specific costs similar to the iceberg cost of trade, which we denote as $\lambda_{ij} \geq 1$. The costs of migration enclose not only the information/monetary/financial costs of moving but also the various policy barriers that deter migration as discussed in Section 2. If one chooses to stay in her original location, there is no additional cost; hence $\lambda_{jj} = 1$ for all location j.

Let J_c denote the set of locations within country c. Combining the three components specified above, an individual living in location j will migrate to i if and only if living in i provides him with the highest utility among all locations within country c:

$$\frac{\bar{v}_i \cdot \iota_i}{\lambda_{ij}} \geq \frac{\bar{v}_{i'} \cdot \iota_{i'}}{\lambda_{i'j}}, \ \ \forall\, i' \in J_c.$$

3.7 Equilibrium

Let $p_{ij}(.)$ and $q_{ij}(.)$ denote the profiles of price and total quantity of differentiated goods sold from j to i across Ω_{ij}, respectively. Let N_j and I_j denote the numbers of workers and entrants in location j, respectively. Let X_i denote the aggregate expenditure on the differentiated goods in location i. Balanced trade implies that X_i is also the total revenue of differentiated goods produced in location i, which is equal to the total costs under the free-entry condition (zero expected profit). As all costs, fixed or variable, are in terms of input bundles specified in (3), the total expenditure on the intermediate goods in location i is $(1-\beta)X_i$. We concisely describe the equilibrium conditions here and refer the reader to Online Appendix B for more details.

Definition: An equilibrium consists of a tuple of prices $\{w_j, p_{ij}(.), P_j^A, R_j(.)\}_{i,j}$, a tuple of quantities $\{N_j, I_j, q_{ij}(.), y_{ji}(.), y_{ji}(,;.), y_j^A, \ell_j(.), \bar{z}_j\}$ for each location i and each urban location j, and a tuple of quantities $\{N_j, q_j^A, L_j, y_{ji}(.), y_j^A, \ell_j\}$ for each location i and each rural location j such that the following conditions hold:

(a) Individuals maximize their utility by choosing locations (including the places of residence within a city if the location is an urban region), residential land consumption, and the consumption bundles from both sectors.

(b) Each firm maximizes its profits by choosing which markets to sell to and the prices charged to each market.

(c) The free-entry condition holds in each location.

(d) The agriculture market clears in each location.

(e) Within each urban region j, land rent $R_j(.)$ clears the urban land market so that urban residents are indifferent across places of residence, $v_j(.) = \bar{v}_j$ and that the city edge $\bar{z}_j$ is such that $R_j(\bar{z}_j) = \bar{R}_j$.

(f) In any rural region j, the land rent R_j clears the land market so that the aggregate land demand $L_j + N_j \ell_j$ equals the total land endowment there.

(g) The differentiated goods market clears such that the aggregate expenditure on the differentiated goods in location i equals the final consumption $\gamma(w_i + T_c)N_i$ and intermediate goods use $(1-\beta) X_i$: $X_i = \gamma(w_i + T_c)N_i + (1-\beta) X_i$.

(h) Labor market clearing for each country c: $\sum_{j \in J_c} N_j = \bar{N}_c$.

3.8 Analytical Solutions

We sketch the analytical solution in this subsection, and refer the reader to Online Appendix B for details.

3.8.1 Firm's Problem

Demand Maximizing the utility function specified in equation (1) yields the demand function faced by firm k located in location j when selling to location i:

$$q_{ij}(k) = \frac{X_i}{(P_i)^{1-\varepsilon}} [p_{ij}(k)]^{-\varepsilon}. \tag{10}$$

The firm takes the aggregate variables X_i and P_i as given when deciding its price, $p_{ij}(k)$. As usual, higher total expenditure and lower firm-level price lead to higher demand. Also, the substitution effect implies that a higher price index in the market (P_i) also increases the demand for firm k as $\varepsilon > 1$.

We solve the firm's problem by backward induction: we start with the pricing decisions conditional on the firm selling to market i; we then outline the decision to sell to market i conditional on firm entry; lastly, we turn to the entry and exit decisions.

Price and Profit in Location i If firm k from j sells to i, the price, $p_{ij}(k)$, is the solution to the profit maximization problem:

$$\pi_{ij}(a) \equiv \max_{p_{ij}(k)} p_{ij}(k)\, q_{ij}(k) - a(k)\, q_{ij}(k)\, \tau_{ij}\chi_j,$$

where $q_{ij}(k)$ is given by (10), and χ_j is the cost of an input bundle in the differentiated sector at location j, which itself is the solution of a cost minimization problem:

$$\chi_j = (w_j)^{\beta} (P_j)^{1-\beta}.$$

Standard procedure yields the constant-markup pricing:

$$p_{ij}(k) = \frac{\varepsilon}{\varepsilon - 1} \tau_{ij}\chi_j a(k).$$

A more productive firm with a lower $a(k)$ is able to charge a lower price, and thus enjoys a larger revenue in region i as the demand elasticity ε is greater than 1. The variable profit is also higher for firms with lower $a(k)$ as π_{ij} is proportional to $(a(k))^{1-\varepsilon}$:

$$\pi_{ij}(a) = \frac{1}{\varepsilon}\frac{X_i}{(P_i)^{1-\varepsilon}}\left(\frac{\varepsilon}{\varepsilon-1}\tau_{ij}\chi_j a(k)\right)^{1-\varepsilon}.$$

At market i, if the market size X_i is larger, or the other firms in the market are relatively unproductive and charge higher prices so that P_i is higher, the variable profit for firm k is higher.

"Exporting" and Total Profit Conditional on the solution of the pricing problem, a firm with input bundle requirement $a(k)$ in location j serves location i if and only if the variable profit covers the fixed cost of trade, f_{ij}: $\pi_{ij}(a) \geq f_{ij}\chi_j$. Moreover, the inequality implies a cutoff rule: the firm in j sells to i if and only if its $a(k)$ is less than a_{ij}:

$$a_{ij} = \frac{\varepsilon-1}{\varepsilon}\frac{P_i}{\tau_{ij}\chi_j}\left(\frac{X_i}{\varepsilon\chi_j f_{ij}}\right)^{\frac{1}{\varepsilon-1}}.$$

A firm in j compares its input bundle requirement to all the cutoffs $a_{ij}, i = 1, 2, \cdots, J$ to determine the market(s) to sell to.

The sales decisions at this stage imply that the total profit of the firm with unit cost $a(k)$, net of the entry costs, is the summation over all the potential markets i:

$$\Pi_j(a(k)) = \sum_{i=1}^{J}\mathbf{1}\left(a(k) < a_{ij}\right)\left(\pi_{ij}(a) - \chi_j f_{ij}\right),$$

where $\mathbf{1}\left(a(k) < a_{ij}\right)$ is an indicator function that equals 1 if the draw is low enough to serve i, and 0 otherwise. More productive firms sell to more markets and earn a higher total profit.

Entry Decision At this final stage, we characterize the entry decision of the potential firms. Prior to paying the entry cost f_e and draw $a(k)$, the *expected profit* of a potential entrant in location j is

$$\overline{\Pi}_j \equiv E\left[\Pi_j(a(k))\right] = \int_0^{1/\mu_j}\Pi_j(a)dG_j(a).$$

The expectation is taken over the distribution of $a(k)$ as characterized by $G_j(a)$. In equilibrium, the expected profit in location j must be equal to the entry costs:

$$\overline{\Pi}_j = f_e \chi_j. \tag{11}$$

Finally, the ideal price index of the differentiated sector in i is the aggregation over all the varieties sourced from all the locations (including itself) as indexed by j:

$$P_i = \left[\sum_{j=1}^{J} \left(\frac{\varepsilon}{\varepsilon - 1} \tau_{ij} \chi_j \right)^{1-\varepsilon} I_j \int_0^{a_{ij}} a^{1-\varepsilon} dG_j\left(a\right) \right]^{\frac{1}{1-\varepsilon}},$$

where I_j is the number of firms that enter the differentiated sector in location j and a_{ij} is the cutoff below which the firm in location j sells to location i. The "love of variety" effect is reflected in the above expression: if more firms are able to sell to market i through either a higher number of entrants (I_j) or a higher cutoff (a_{ij}), then the ideal price index in location i is lower.

3.8.2 Equilibrium within a City and Aggregate Land Rent

For an urban region i, let $\hat{N}_i$ be the average city size within the urban region so that N_i is the product of $\hat{N}_i$ and the number of cities within the region, which is exogenously given. As N_i is determined in cross-region spatial equilibrium, we take $\hat{N}_i$ as given when considering the equilibrium within the city.

The income of an individual in an urban region i in country c is given by $w_i + T_c$, and the Cobb-Douglas structure entails the indirect utility (8). Following Ahlfeldt et al. (2015), we let $d\left(z\right) = e^{\delta \tau(z)}$, where $\tau\left(z\right)$ is travel time from z to the CBD. Furthermore, assume that $\tau\left(z\right) = tz$, namely, travel time is linear in distance. The indifference condition $v_i(.) = \bar{v}_i$ implies that

$$R_i(z) = \left(\frac{w_i + T_c}{e^{\delta t z} \left(P_i^A\right)^{\alpha} P_i^{\gamma} \bar{v}_i} \right)^{\frac{1}{1-\alpha-\gamma}}. \tag{12}$$

The population density at each point z is the inverse of land use per person, i.e., $1/\ell_i(z)$. The land market clearing condition is therefore[9]

$$\hat{N}_i = \int_0^{\bar{z}_i} 2\pi z \frac{1}{\ell_i(z)} dz. \tag{13}$$

Using (12), (13), and the facts that $\ell_i(z) = (1-\alpha-\gamma)(w_i+T_c)/R_i(z)$ and that $R_i(\bar{z}_i) = \bar{R}_i$, it is readily obtained that

$$e^{\frac{\delta t \bar{z}_i}{1-\alpha-\gamma}} - \left(1 + \frac{\delta t \bar{z}_i}{1-\alpha-\gamma}\right) = \frac{\delta^2 t^2 (w_i + T_c) \hat{N}_i}{2\pi (1-\alpha-\gamma) \bar{R}_i}. \tag{14}$$

Taking $(w_i + T_c)$ and $\hat{N}_i$ as given, this is the single equation that pins down the city radius $\bar{z}_i$ and summarizes the information on commuting costs and land rents. Observing the indirect utility (8) and noting that $v_i(.) = \bar{v}_i$, we see that $d(z)R_i(z)^{1-\alpha-\gamma} = C$ for some constant C. Using $R_i(\bar{z}_i) = \bar{R}_i$, one can solve out this constant C and arrive at the equilibrium land rent:

$$R_i(z) = \bar{R}_i e^{\frac{\delta t (\bar{z}_i - z)}{1-\alpha-\gamma}}. \tag{15}$$

The equilibrium utility is therefore given by

$$\bar{v}_i = \frac{\phi_i (w_i + T_c)}{e^{\delta t \bar{z}_i} (P_i^A)^\alpha P_i^\gamma \bar{R}_i^{1-\alpha-\gamma}}. \tag{16}$$

Next, we consider how land rent rebate T_c is determined; for this we must derive the aggregate land rent. The Cobb-Douglas structure implies that the aggregate land rents in an urban region i and the aggregate residential land rents in a rural region i are both given by $(1-\alpha-\gamma)(w_i+T_c)N_i$. Recall that J_c denotes the set of locations in country c, and further define $R_{A,c}$ and $L_{A,c}$ as agricultural land rent and land use in country c, respectively.[10] We have

$$T_c = \frac{R_{A,c}L_{A,c} + (1-\alpha-\gamma)\sum_{i\in J_c}(w_i+T_c)N_i}{\bar{N}_c},$$

[9]As in a standard monocentric city model, the land-market-clearing condition is conveniently expressed as the integral of population density over the city space (the disk with a radius $\bar{z}_i$) which should equal to the total city population. This is because the population density is the inverse of land demand per person ($1/\ell_i(z)$), whereas the area of the disk gives the land supply.

[10]Recall that the rural land rent R_i for land consumption is equal to $R_{A,c}$ when the market clears.

which entails

$$T_c = \frac{R_{A,c}L_{A,c} + (1-\alpha-\gamma)\sum_{i\in J_c} w_i N_i}{(\alpha+\gamma)\bar{N}_c}.$$

The right-hand side of the above equation implicitly depends on $\{T_c\}$ across countries via the general equilibrium objects in the numerator. We solve for $\{T_c\}$ using the above equation as a fixed-point mapping.

3.8.3 Migration Decision

It is straightforward to show that conditional on $\{\bar{v}_i\}_{i\in J_c}$, the fraction of the population that migrates from j to i in country c is

$$m_{ij} = \frac{(\bar{v}_i)^{\kappa}(\lambda_{ij})^{-\kappa}}{\sum_{i'\in J_c}(\bar{v}_{i'})^{\kappa}(\lambda_{i'j})^{-\kappa}}. \tag{17}$$

The above equation is similar to that used in Redding (2016) and Tombe and Zhu (2019) and is related to the "gravity equation" in international migration flows such as those in Grogger and Hanson (2011) and Ortega and Peri (2013). Moreover, note that κ is the migration elasticity with respect to friction. The larger the κ, the less heterogeneous the idiosyncratic locational preferences, and hence the more sensitive migration flows are to changes in migration friction.

4 Quantification

To quantify the model, we group the world into three countries: China (CHN), other developing countries (ODC), and the rest of the world (ROW). As the calibration strategy requires data from the World Development Indicators (WDI), the Penn World Table (PWT), and the Inter-Country Input-Output (ICIO) tables, we take the largest intersection of countries from these three datasets as our sample of countries. Out of the 63 countries in the sample, we group the countries whose average per capital GDP is less then 2/3 of the USA's into the "other developing countries", and the rest as the ROW. Online Appendix C provides the details about the sample definition, and Online Appendix Table A.2 lists all the countries in the sample. As mentioned in Section 2.2, China is divided into three regions: rural ($r = 1$), the MUR ($r = 2$), and the OUR ($r = 3$). The other two countries only contain one rural and one urban region each.[11]

We calibrate the model to the world economy around year 2005 and around year 2015 separately. The model parameters fall into one of the two categories. The first group of parameters is common across the two years, and the second group is calibrated to each year. Within the year-specific parameters, some parameters are calibrated directly from the data without solving the model, while the others are jointly calibrated based on model simulations. In this section, we introduce the calibration strategy for each of the parameters, and refer the reader to Online Appendix C for more details. All parameters are summarized in Table 2 and 3.

4.1 Common Parameters

This group of parameters is common across the two years. Table 2 summarizes these parameters:

- The labor share in differentiated products, $\beta = 0.37$. The data source is the Input-Output Table in China in 2002. This parameter is the ratio between the total value-added and the total output across all non-agriculture industries. This calibration strategy compensates for the absence of capital in the production function by treating the return to capital as part of the return to labor.

- The elasticity of substitution, ε, and the Pareto tail index, θ. As the tail index also serves as the trade elasticity in the model, we set $\theta = 4$ following the estimates in Simonovska and Waugh (2014). Moreover, $\frac{\theta}{\varepsilon-1}$ equals the tail index of the employment distribution of firms in equilibrium. In light of this, we set $\varepsilon = 4.717$ so that the tail index equals 1.076, the value reported in Ma and Tang (2020b) based on the Chinese plant-level data.[12]

- The shape parameter of the distribution of locational preference, $\kappa = 1.63$. This parameter is also the elasticity of migration flows with respect to friction. Monte et al. (2018), using the same Fréchet distribution, estimate this parameter to be 3.3 in the context of the US and Hsieh and Moretti (2019) set it to 2.0 based on a similar extreme-value distribution. Bryan

[11] Note that as we lump countries besides China into two blocks, it is difficult to interpret the rural-urban migration within each block. As our focus is on China's patterns of rural-urban migration, the rural and urban population in these two country blocks are held fixed in our quantitative exercises.

[12] Note that this value is rather close to 1.06, the value reported by Axtell (2001) and Luttmer (2007) using plant-level data from the US.

Name	Value	Source	Note
β	0.37	Input-Output Table, 2002	Labor share in differentiated goods production
ε	4.717	Firm size distribution in China	Elasticity of substitution
θ	4.0	Simonovska and Waugh (2014)	Trade elasticity and Pareto tail index in productivity distribution
κ	1.63	Ma and Tang (2020a)	Migration elasticity and shape parameter in location preference
δ	0.01	Ahlfeldt et al. (2015)	Semi-elasticity of iceberg commuting cost to travel time (minutes)
t	4.35	Baidu Hundred-City Commuting Data	Travel time in minutes per unit distance
α	0.15	Input-Output Table, 2005	Expenditure share of agricultural goods in 2005
	0.07	Input-Output Table, 2015	Expenditure share of agricultural goods in 2015
γ	0.7875	$1-\alpha-0.0625$ (see Section 4.2.1)	Expenditure share of the differentiated products in 2005
	0.8675	$1-\alpha-0.0625$ (see Section 4.2.1)	Expenditure share of the differentiated products in 2015

Table 2: Parameters

Note: This table summarizes the calibrated model parameters common across years and the year-specific α and γ.

and Morten (2019) estimate it to be 2.7 using Indonesia data. The results from the reduced form gravity-equation estimations often suggest that the distance elasticity of migration is generally smaller than 2.0. In the case of the European countries, it is found to be around between 1.4 and 2.2 in (Stillwell et al., 2014). In this paper, we use $\kappa = 1.63$ based on Ma and Tang (2020a), which is estimated from the *One-Percent Population Survey* in 2005. We will later use $\kappa = 3.3$, an estimate from the higher end of the spectrum, as a robustness check in Section 5.5. As will be seen in Section 5.5, all the main welfare results are strengthened in a world with a higher migration elasticity as the population movements are more sensitive to changes in urbanization policy. In this sense, the baseline results reported in this paper are conservative estimates of the impact of urbanization policies.

- The commuting costs. Following Ahlfeldt et al. (2015), the semi-elasticity of iceberg commuting cost to travel time (minutes), δ, is set to 0.01. As the commuting time is assumed to be a linear function of distance, tz, the parameter t, i.e., the travel time per unit distance, is set to 4.35, which is the average travel time in minutes per kilometer across all cities in the Baidu Hundred-City Commuting Cost Report.

4.2 Year-Specific Parameters

All of the other parameters in the model are calibrated to match the data moments in year $t = 2005$ and 2015 separately. We first introduce the parameters that are calibrated without solving the model, and then move to the simulation-based joint calibration.

4.2.1 Expenditure Shares, Initial Population, Trade Cost, Urban Productivity, Land Rent, and Agricultural Production

Expenditure Shares The expenditure share of agriculture products, α, comes from the Input-Output Table from China in the year 2005 and 2015 respectively. We back out this parameter using the household expenditure on agriculture products as a share of total consumption. In 2005, $\alpha = 0.15$, and in 2015, $\alpha = 0.07$. We set the expenditure share of land consumption to 0.0625, which is the product of housing expenditure share (0.25 from Davis and Ortalo-Magné, 2011) and the share of land in the housing construction costs (0.25 from Combes et al., 2019). The expenditure share of the differentiated products, γ, is then inferred by $\gamma = 1-\alpha-0.0625 = 0.7875$ in 2005 and $\gamma = 0.8675$ in 2015.

Initial Population To construct the initial population distributions for the 2005 and 2015 models, our starting point is the country-level population data in the year 2000 and 2010 from the Penn World Table, respectively. Multiplying the total population by the percentage of the workforce employed in agriculture from the WDI yields the rural population in each country. The total urban population is simply the difference between the total and rural populations. Within China, the total urban population must be further divided between the MUR and OUR. To do this, we use the distribution of prefecture-level urban population, i.e., the population in the collection of "districts" within a prefecture (*Shi Xia Qu*), from the 2000 and 2010 Population Censuses to allocate the total urban population to the two regions. In the last step, we normalize the initial population headcount in both periods so that the rural population in China in the year 2000 is 1.0.

Trade Costs We follow di Giovanni and Levchenko (2012) by using the Doing Business Database from the World Bank to estimate the fixed costs of trade. We take the number of days required to start a new business in each country in year $t = \{2005, 2015\}$ and compute the population-weighted average within each country group to estimate f_{ii}, the fixed costs of country i to sell to its own market. We then construct the fixed costs of exporting from j to i as $f_{ij} = f_{ii} + f_{jj}$. Lastly, we normalize all the fixed costs so that f_{ii} for China equals 1. Conditional on the estimated fixed costs of exports, we then follow Novy (2013) to back out the variable trade costs, τ_{ij}, from the observed trade flow. The trade flow data, which includes domestic absorption, come from the ICIO Tables provided by the OECD. We assume free trade between regions within each country. In our context, free trade means that $\tau_{ij} = 1.0$ between all i and j within the same country so trade does not suffer from iceberg friction, and $f_{ij} = f_{jj}$ so that firms originating from region j face the same fixed costs in selling locally and to another region i in the same country.[13]

We estimate the variable trade costs for agricultural goods as proportional to τ_{ij}, so that $\tau_{ij}^A = \bar{\tau} \times \tau_{ij}$. The ESCAP-World Bank Trade Cost Database provides estimates of the variable trade costs by country and industry, which includes agriculture, following the methods in Novy (2013). We take the average ratio of agriculture trade costs to manufacturing trade costs in that database as the estimate of $\bar{\tau}$ and arrive at $\bar{\tau} = 2.14$.

Productivity in Urban Regions The productivities in the urban regions, $\{\mu_j\}$, are estimated based on the cross-sectional TFP (cTFP) data provided by the PWT. In year $t = \{2005, 2015\}$, we take the population-weighted average within each country group to obtain the estimates for all three countries. As the ODC and ROW only have a single urban region, the above step provides the productivities in these countries up to a scale; however, more work needs to be done to infer the urban productivity in the two urban regions in China.

To back out the productivities in the two urban regions in China, we start by estimating the productivity for each of the 279 cities (which correspond to 279 prefectural-level cities) in China. We follow the estimation strategy outlined in Ma and Tang (2020b), who apply the methods in Donaldson and Hornbeck (2016) to the context of China. We outline the estimation in Online Appendix C and refer readers to Ma and Tang (2020b) for more details. With the estimated productivity for each individual city, it is straightforward to aggregate by taking the population-weighted average within each urban region to arrive at an estimate of the *ratio* of productivity between the two urban regions.

To back out the *level* of the urban productivity, we further require that the population-weighted average of the two urban productivities be equal to the cTFP in China from the PWT. This additional constraint implies that the country-level TFP in China coming from the previous exercise is

[13] As is common in the Melitz model with entry costs (f_e), we scale the $\{f_{ij}\}$ matrices by an arbitrary number to ensure an interior solution of the cut-offs in all the baseline and counter-factual simulations.

	Rural China	MUR	OUR	Rural ODC	Urban ODC	Rural ROW	Urban ROW
N_i, 2005	1.00	0.30	0.76	1.62	2.41	0.04	1.27
N_i, 2015	0.71	0.42	1.05	1.48	3.05	0.03	1.37
μ_i, 2005	1.00	1.00	0.86	1.02	1.14	1.32	2.33
μ_i, 2015	1.31	1.29	1.11	1.22	1.52	1.59	2.63
ϕ_i, 2005	1.04	0.90	1.05	-	-	-	-
ϕ_i, 2015	1.06	1.00	0.99	-	-	-	-
$\bar{R}_i$, 2005	-	124.24	51.65	-	-	-	-
$\bar{R}_i$, 2015	-	657.16	125.15	-	-	-	-
ν_i, 2005	0.33	-	-	0.45	-	0.27	-
ν_i, 2015	0.33	-	-	0.45	-	0.26	-
η_i, 2005	0.26	-	-	0.22	-	0.20	-
η_i, 2015	0.26	-	-	0.20	-	0.24	-
Land, 2005	1.00	-	-	3.16	-	1.62	-
Land, 2015	1.09	-	-	3.30	-	1.52	-

(a) Initial Population, Productivity, Amenity, Outside Land Value, Agriculture Production Function, and Land Endowments

	China (o)	ODC (o)	ROW (o)
China (d)	1.00	2.59	2.55
ODC (d)	2.59	1.00	2.07
ROW (d)	2.55	2.07	1.00

(b) τ_{ij}, 2005

	China (o)	ODC (o)	ROW (o)
China (d)	1.00	2.46	2.48
ODC (d)	2.46	1.00	2.05
ROW (d)	2.48	2.05	1.00

(c) τ_{ij}, 2015

	China (o)	ODC (o)	ROW (o)
China (d)	1.00	2.50	1.80
ODC (d)	2.50	1.50	2.30
ROW (d)	1.80	2.30	0.80

(d) f_{ij}, 2005

	China (o)	ODC (o)	ROW (o)
China (d)	1.00	1.96	1.32
ODC (d)	1.96	0.96	1.28
ROW (d)	1.32	1.28	0.32

(e) f_{ij}, 2015

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	1.00	3.14	2.03
MUR (d)	21.58	1.00	10.02
OUR (d)	20.74	17.24	1.00
f_e	6.98		

(f) Joint Calibration of λ_{ij} and f_e, 2005

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	1.00	1.02	0.67
MUR (d)	15.05	1.00	4.86
OUR (d)	5.68	6.93	1.00
f_e	10.67		

(g) Joint Calibration λ_{ij} and f_e, 2015

Table 3: Year-Specific Parameters

Note: This table summarizes the calibrated model parameters that are year-specific, excluding α and γ that are reported in Table 2.

comparable to the TFP measures for the ODC and ROW. This constraint, together with the relative productivity in the previous step, constitutes a simple two-equation system from which the two urban productivities in China can be computed.

Lastly, we normalize the productivity vectors so that $\mu_j = 1$ in the MUR of China in 2005. To reflect the total productivity growth between the two years, we multiply the entire productivity vector in 2015 by 1.2946, the ratio between the inter-temporal measure of TFP (rTFPna) in China between 2015 and 2005 from the PWT.

As reported in Table 3a, we find that the relative productivity between the two urban regions in China barely moves between the two years. The productivity of the OUR is 86% of the MUR back in 2005, and the relative productivity barely moves ten years later. As there is no "regional convergence" in productivity, the productivity gap between the two urban regions does not explain the reversal of migration patterns observed in Table 1.

Agricultural Land, Production Function, and Productivity The variables related to the rural regions come from the *U.S. Department of Agriculture, Economic Research Services* (USDA-ERS) database. All of these parameters are reported in Table 3a.

We use the total agricultural land, which includes un-irrigated and irrigated cropland and pastures, as the measure of the land endowment of each country's rural region. We sum the agricultural land to the country-group level and normalize the land endowment in China in the year 2005 to 1.

The database provides a break-down of factor input costs into six categories: labor, land, livestock, machinery, fertilizer, and feed. The first two categories directly correspond to labor and land in our model, respectively. We treat the remaining four factors as intermediate inputs in our model. In a given year, we aggregate the land and labor shares to the country-group level by taking weighted averages. The weights are the rural population and the rural land endowments of each country.

The agricultural productivity is then computed as the exponentiated residual by regressing the logarithm of agricultural output on the logarithm of all six inputs at the country level. The country-group-level agricultural productivity is obtained as the weighted average of country-level agricultural productivity with the weights being the agricultural output. Within each year, we normalize the agricultural productivity so that the rural productivity in China is 1. Lastly, similar to the urban productivity, to reflect the total productivity growth between the two years, we multiply the rural productivity vector in 2015 by 1.20, the average growth rate of rural productivity across countries between 2015 and 2005 in USDA-ERS.[14]

Outside Land Rent, $\bar{R}_i$ Equations (14) and (15) offer equilibrium conditions to back out $\bar{R}_i$ in the urban regions in China. We first evaluate Equation (15) at $z = 0$ and use the CBD land rent as proxies to the left-hand side (LHS) of the equation to pin down the relationship between $\bar{R}_i$ and $\bar{z}_i$.[15] In the second step, we then use Equation (14) to back out the outside land rent, $\bar{R}_i$. We proxy the term $(w_i + T_c)\hat{N}_i$ in each region using the GDP of an average city in the respective region.

Table 3 reports the estimates of $\bar{R}_i$ for the two urban regions in China. The rural R_i in China is not reported, as it is endogenously determined from the rural land market. We nevertheless compute a rural R_i from the data as it serves as a reference point to convert the $\bar{R}_i$ in the urban regions in China to the same units as the endogenous land rents in the model.[16] To compute the R_i for the rural regions, we use the inferred rural land rent. We use the factor share data from USDA-ERS and the rural output and land endowment data from the City Statistical Yearbooks to estimate the annual rural land rent as the ratio between the expenditure on land and the land endowment.[17]

The $\bar{R}_i$'s for other countries' urban regions are given by equilibrium rural land rent; thus, it is not an object of calibration. We adopt this approach for two reasons. First, the institutional barriers to rural-urban land transfers in other countries are generally lower than China's; second, there is no clear-cut way of defining the outside land rent in the context of country groups instead of individual countries.

The estimated $\bar{R}_i$'s are reported in Table 3a. The outside land rent is always higher in the MUR than in the OUR, as expected. However, the gap widens drastically from 140% in 2005 to more

[14] In the quantification, we normalize both the productivity in the MUR and China's rural region to 1 in 2005. There is no need to specify the relative productivity between the rural region and the MUR in China, as it does not affect the equilibrium outcome, including the migration and trade patterns. This is because the Cobb-Douglas utility function implies the irrelevance of the relative productivity between the two sectors. Intuitively, scaling all of the rural productivities by a constant affects only the price and not the agriculture product's expenditure share.

[15] See Online Appendix C for how CBD land rents are calculated.

[16] The $\bar{R}_i$ for the urban regions in China as computed in the previous steps are denoted in Chinese RMB and are thus not consistent with the endogenously determined land rent in the model which uses rural wage rate as the numeraire. We use the R_i computed here for rural China and the endogenously determined land rent in rural China to compute the exchange rate between the numeraire in the model and RMB and convert the urban $\bar{R}_i$ from the data to the model units.

[17] The land endowment data from USDA-ERS cannot be used to infer land rent as the unit of land is in "efficiency units", not physical units (hectares) as in the CBD land rent data.

than 500% in 2015.

4.2.2 Joint Calibration

The above quantification procedure leaves 10 parameters to be jointly calibrated in each year. These parameters are the entry cost, f_e, the amenity for the three regions in China, and the six off-diagonal terms in the migration cost matrix in China, λ_{ij}. We pin down these parameters using an iterative procedure as follows.

Step 1: With an initial guess of $\{\lambda_{ij}\}$, we use Equations (16–17) and the data on the bilateral migration flow and city-level characteristics to estimate $\{\phi_i\}$.

Step 2: Conditional on $\{\phi_i\}$, we solve the model to jointly calibrate the remaining 7 parameters (f_e and $\{\lambda_{ij}\}$) using 7 moments in the data.

Step 3: Go back to Step 1 until $\{\lambda_{ij}\}$ converge.

In practice, we start with the initial guess of $\lambda_{ij} = 1, \forall i, j$ and use 1.0E-4 as the threshold for convergence. The procedure converges in less than 5 iterations. In the rest of this section, we describe the details of each step. These parameters are reported in Tables 3a, 3f, and 3g.

Step 1: Estimating Amenity Conditional on λ_{ij} Conditional on a guess of λ_{ij}, we use the observed between-city migration flow and the city-level characteristics to backout the amenity of the urban regions in three steps. The overall strategy is similar to the estimation of productivity as explained above: we first estimate the amenity at the city level, and then aggregate up to the region level. Note that as we only study the migration problem in China, we abstract away from the amenities in the other two countries.

In the first step, we interpret Equation (17) at the city level and express bilateral migration flows as

$$\log\left(m_{ij}N_j\right) = \kappa \log(\bar{v}_i) + \log(N_j) - \log\left(\sum_{i' \in J_c} (\bar{v}_{i'})^{\kappa} \left(\lambda_{i'j}\right)^{-\kappa}\right) - \kappa \log(\lambda_{ij}). \tag{18}$$

In equation 18, the LHS is the observed migration flow from j to i. We regress the LHS against the destination and origin fixed effects, as well as the $\log(\lambda_{ij})$ from the previous iteration. In this regression, the destination fixed effect of city i, which we denote as $\mathbf{D}_i$, absorbs the indirect utility of city i as $\mathbf{D}_i = \kappa \log(\bar{v}_i)$.[18]

For the second step, first observe that Equation (15) implies that $\delta t \bar{z}_i = (1-\alpha-\gamma)\left[\log(R_i(0)) - \log(\bar{R}_i)\right]$. Combining this with Equation (16) entails

$$\log(\bar{v}_i) = \log(\phi_i) - (1-\alpha-\gamma)\log(R_i(0)) + \log\left[\frac{w_i + T_c}{(P_i^A)^\alpha P_i^\gamma}\right].$$

We then combine this expression with the earlier expression of the destination fixed effects:

$$\frac{1}{\kappa}\mathbf{D}_i = -(1-\alpha-\gamma)\log(R_i(0)) + \log(w_i + T_c) - \log\left[\left(P_i^A\right)^\alpha (P_i)^\gamma\right] + \log(\phi_i). \tag{19}$$

We regress the $\frac{1}{\kappa}\mathbf{D}_i$ from the first step on $\log(R_i(0))$, approximated by the CBD land rent, and per capita GDP to capture the second term. Therefore, the residual from this regressing equation 19 contains the logarithm of amenity, as well as that of prices.

In the last step, we linearly project the residual, denoted as $\mathbf{e}_i$, on a vector of city-level characteristics, $\mathbf{X}_i$, that is related to amenity:[19]

$$\mathbf{e}_i = \mathbf{b}_0 + \mathbf{b}_1 \mathbf{X}_i,$$

and use the prediction from the projection as the estimate for $\log(\phi_i) = \widehat{\mathbf{b}}_1 \mathbf{X}_i$ at the city level. This last step relates the estimated amenity to a wide range of observed characteristics. Moreover, by capturing these elements in ϕ_i, the projection ensures that the estimated λ_{ij} from the next stage will not be contaminated by the same elements.

With the city-level amenity, we aggregate the amenity at the city level to the region level, using the city population as weights. Note that in the entire procedure, we treat the urban and rural areas of a prefecture as separate locations; in total we thus have $279 * 2$ "cities" in each regression.

[18]We note that this step also yields an estimate of κ. We prefer to use the κ from the literature as explained in the earlier parts of this section since the estimation procedure outlined here is not designed to estimate the migration elasticity. We control for λ_{ij} to ensure the consistency of the estimated $\mathbf{D}_i$, not to estimate κ. The same considerations apply to the estimated $1-\alpha-\gamma$ as well: we prefer to use the α and γ reported in Table 2.

[19]City-level characteristics include the average temperature, precipitation, elevation, and slope; the number of universities, middle schools, and primary schools; the number of university, middle school, and primary school teachers; the number of public library books; the number of hospitals, hospital beds, and doctors; the percentage of green fields in constructed areas; and the ease of access to transportation networks. More details are provided in Online Appendix C.

Step 2: Calibrating f_e and λ_{ij} Conditional on ϕ_i Conditional on ϕ_i from the previous step, we then solve the model and calibrate the remaining 7 parameters by targeting 7 equilibrium moment conditions in the model.

The entry cost f_e determines the numbers of entrants in urban regions. We use this parameter to match the firms-to-population ratio in the MUR in China. The number of firms data come from the 2004 and 2014 Economic Census in China for the quantification in 2005 and 2015, respectively, and our target moment is 15.1 and 19.3 entering firms per thousand population.

The last six λ_{ij} parameters are pinned down by the migration probability matrix estimated from the *One-Percent Population Survey* as presented in Table 1. The migration friction $\{\lambda_{ij}\}$ is backed out by matching the model migration probability $\{m_{ij}\}$ given in Equation (17) to the observed migration probability $\{\tilde{m}_{ij}\}$ in Table 1.

4.3 Discussion on Parameter Estimates and Possible Explanations

In this subsection we make several observations based on parameter estimates. First, the migration costs, $\{\lambda_{ij}\}$, decline substantially between 2005 and 2015: the average magnitude drops by 65% within this 10-year span. The relaxation of migration friction is strongest among the rural-to-urban flows: the costs of moving from the rural region to the MUR drop by 56%, and to the OUR, 62%. The changes in the other bilateral migration costs are lower and yet still sizable. The sharp decline in the calibrated λ_{ij} parameters is underpinned by the significant shifts in the migration probability matrix in the data as presented in Table 1. For example, in 2005 around 94% of the individuals originating from rural areas choose to stay in rural areas, whereas 10 years later, the same statistic plunges to only 62%.

Comparing the migration friction across the two periods also reveals a fundamental change in the urbanization policy that favors smaller cities. In 2005, the costs of moving from the rural region to the two urban regions in China are roughly the same at $\lambda_{21} = 21.58$ and $\lambda_{31} = 20.74$. However, 10 years later, it is significantly easier to move to the OUR ($\lambda_{31} = 5.68$) than to the MUR ($\lambda_{21} = 15.05$) from the rural area. The migration probability from the data reveals the same pattern: the rural population is equally likely to move to the two urban regions in 2005; fast-forward to 2015, and the pattern shifts and the rural population is 250% more likely to migrate to the OUR.[20]

We note that λ_{ij} is not directly observed in the data but rather inferred through the lens of our model. However, our model does not assume that migration frictions are the most important driving forces behind the observed migration pattern *a priori*. Instead, we include competing forces such as productivity, amenity, and the urban costs in the model. Which forces explain the observed pattern should be determined by the data and counter-factual analysis based on the calibrated model, as we do in this and the next sections.

The drastic change in migration flows is unlikely to be driven by regional productivity, as we see little regional convergence in productivity. Moreover, amenity in the MUR grew faster than that in the OUR between the two periods, as seen in Table 3, implying that amenity alone would have predicted the opposite pattern of that seen in 2015. Nevertheless, we also carry out counter-factual analyses on productivity and amenity and confirm that these two factors have little explanatory power on the observed migration pattern. The details are relegated to Online Appendix E. Therefore, the contrast between the migration patterns in these two periods is likely to be explained by changes in either migration friction or urban land markets (which are affected by land supply), or both. We will further evaluate the relative importance of these two channels in the next section.

5 Quantitative Results

In this section, we carry out several counter-factual exercises to shed light on the impact of urbanization policy in China.

5.1 Impact of Differential Reform on Migration Restrictions

We first evaluate the impact of differential reform on migration restrictions by considering two alternative urbanization policies. As discussed above, the baseline calibration in the year 2015 captures a basic migration pattern that is consistent with urbanization policy pivoting towards small- and medium-sized cities. In the model, the emphasis towards smaller cities is reflected in the λ matrix reported in Table 3: whereas migration costs from the rural region to the OUR drop from 20.74 in 2005 to 5.68 in 2015, rural-MUR migration costs decline more mildly from 21.58 to

[20] From Table 1, $0.294/0.084 - 1 \approx 250\%$.

15.05. As a result, during 2005–2015, out of the 37.8% of the rural population that emigrates, the majority, 29.4%, goes to the smaller cities; only 8.4% chooses to move to large cities.

To evaluate the impact of such uneven reduction in migration costs, we simulate two counter-factual cases. In the first exercise, hereafter referred to as the "λ^* counter-factual", we eliminate the pivot towards the OUR by equalizing the rural-urban migration costs across the two urban regions, e.g., setting $\lambda_{31} = \lambda_{21} = \lambda^*$. We pick the value of λ^* so that the same 37.8% of the rural population chooses to move out. In other words, we simulate a world in which the same number of people move into the two urban areas facing the same migration friction. In practice, we find $\lambda^* = 8.35$; as a result, the inbound friction towards the large cities is relaxed, while the friction into the small cities is tightened. In the second counter-factual analysis, referred to as the "low λ counter-factual", we equalize the two rural-urban migration friction to the lower value of the two, so that $\lambda_{31} = \lambda_{21} = \min\{\lambda_{31}, \lambda_{21}\} = 5.68$. The rationale behind this exercise is to extend the more liberal migration policy to the rural-MUR migration as well. Note that as this exercise is an overall relaxation of migration friction, we expect a higher volume of rural emigration and larger overall welfare gain as well. In both counter-factual exercises, we simulate the model using all of the other parameters from the 2015 quantification. The results are reported in the second and third panels of Table 4, where the first panel presents the baseline in 2015 for comparison.

5.1.1 λ^* counter-factual

We first focus on the λ^* counter-factual in which the same number of people move out of the rural region as compared to the baseline. Under the equalized migration costs, rural migrants are more likely to move to the MUR, as reported in Panel (b) of Table 5. Unlike the baseline case in Table 1, in the counter-factual simulation 22.1% of the rural population are now moving into large cities, and 15.6% into smaller cities. This reverses the migration pattern empirically observed as shown in Table 1(b). The re-directed migration flow is approximately 4.4% of the total population.[21] The popularity of the large cities is expected as it enjoys a higher productivity and amenity. As a result of the re-directed migration flow, the population in the MUR increases by $0.6209/0.5209 - 1 \approx 19.2\%$ whereas the population in the OUR drops by around 10.0%.

[21] Between the counter-factual and the baseline, $22.1 - 8.4 = 13.7\%$ of the rural emigrants are re-directed towards the MUR. As the initial population in rural China is 0.7080, the re-directed flow is $0.137 * 0.7080/2.1847 = 4.4\%$.

	China	Rural	MUR	OUR	Rural ODC	Urban ODC	Rural ROW	Urban ROW
	Baseline 2015							
Population	2.1847	0.6629	0.5209	1.0009	1.4761	3.0510	0.0322	1.3668
Welfare	0.2256	0.0609	0.2823	0.2347	0.2747	2.1750	6.8808	10.2857
Operating Firms	2.4687	-	1.2384	1.2303	-	190.8005	-	6231.7030
Exporting Firms	0.4261	-	0.2137	0.2123	-	11.6143	-	54.0165
	$\lambda_{21} = \lambda_{31} = \lambda^*$							
Population	2.1847	0.6627	0.6209	0.9010	1.4761	3.0510	0.0322	1.3668
Welfare	0.2314	0.0620	0.2901	0.2380	0.2745	2.1734	6.9098	10.2743
Operating Firms	2.2755	-	1.3047	0.9708	-	185.2347	-	6051.6378
Exporting Firms	0.4376	-	0.2509	0.1867	-	11.2766	-	52.4507
	Low λ							
Population	2.1847	0.5766	0.6763	0.9317	1.4761	3.0510	0.0322	1.3668
Welfare	0.2608	0.0684	0.3081	0.2522	0.2749	2.1766	7.2183	10.2764
Operating Firms	1.6474	-	0.9658	0.6815	-	119.0079	-	3901.1131
Exporting Firms	0.3071	-	0.1800	0.1270	-	7.2535	-	33.7712
	Low Growth of $\bar{R}_{\text{MUR}}$							
Population	2.1847	0.6577	0.5348	0.9922	1.4761	3.0510	0.0322	1.3668
Welfare	0.2315	0.0614	0.2982	0.2357	0.2747	2.1752	6.9009	10.2854
Operating Firms	2.4075	-	1.2290	1.1784	-	185.5566	-	6061.7794
Exporting Firms	0.4191	-	0.2139	0.2051	-	11.2960	-	52.5395

Table 4: Impact of Urbanization Policies

Note: This table lists the key endogenous variables for all 7 regions across the baseline and the counter-factual simulations. The first column is the aggregate result for China.

The national welfare is defined as the average of the region-level welfare, $\bar{v}_i m_{ii}^{-1/\kappa}$, across the regions in the country, weighted by the equilibrium regional population. Recall that $\bar{v}_i$ is given in Equation (16). The $m_{ii}^{-1/\kappa}$ term appears as it captures the welfare loss due to migration frictions; see Proposition 2 in Tombe and Zhu (2019).

The national welfare increases by $0.2314/0.2256 - 1 \approx 2.6\%$ by adopting the λ^* policy. The aggregate gain in welfare is substantial, considering the relatively modest change in migration flow of 4.4% of the total population. The welfare also increases in all three regions in China. The MUR enjoys a higher welfare due to the increased population base and market size, which in turn supports a higher number of operating firms and higher average productivity due to fiercer selection. Even with a reduced population, the OUR still enjoys a higher welfare due to the spillover effect from large cities. As labor supply concentrates in the high-productivity area, both intermediate

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.622	0.078	0.180
MUR (d)	0.084	0.894	0.079
OUR (d)	0.294	0.027	0.742

(a) Baseline 2015

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.622	0.077	0.180
MUR (d)	0.221	0.896	0.080
OUR (d)	0.156	0.027	0.740

(b) λ^*

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.484	0.082	0.190
MUR (d)	0.303	0.892	0.079
OUR (d)	0.213	0.027	0.731

(c) Low λ

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.619	0.073	0.179
MUR (d)	0.090	0.901	0.085
OUR (d)	0.291	0.026	0.737

(d) Low growth of $\bar{R}_{\text{MUR}}$

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.629	0.080	0.183
MUR (d)	0.082	0.892	0.078
OUR (d)	0.289	0.027	0.738

(e) Welfare-equivalent trade liberalization, λ^*

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.645	0.086	0.194
MUR (d)	0.078	0.887	0.077
OUR (d)	0.276	0.027	0.729

(f) Welfare-equivalent trade liberalization, low λ

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.622	0.079	0.184
MUR (d)	0.222	0.894	0.080
OUR (d)	0.155	0.027	0.736

(g) λ^*, autarky

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.507	0.089	0.205
MUR (d)	0.291	0.884	0.078
OUR (d)	0.203	0.027	0.718

(h) low λ, autarky

Table 5: Migration Probability Matrices in Counter-factual Simulations

Note: This table presents the migration probability matrix within China in various model simulations. An element at the i-th row and the j-th column indicates the probability of an individual originating from j and moving to i. Each column sums to 1.

and final consumption goods coming out of large cities become cheaper, which in turn benefits the other two regions through intra-national trade.

The changes in urbanization policy also lead to ramifications around the world. After China adopts the λ^* policy, the population concentration in large cities lowers labor costs there. Chinese exporters become more competitive in overseas markets and subsequently drive out inefficient firms in these countries. As a result of this higher import competition, the number of operating firms in ODC and ROW drops by around 2.9%.[22] Similarly, the improved aggregate productivity in China means that the Chinese market is harder to export to as well; correspondingly, the num-

[22] $1 - 185.2/190.8 \approx 2.9\%$ in the ODC, and $1 - 6051/6231 \approx 2.9\%$ in the ROW.

bers of exporters in the ODC and ROW also decline by around the same magnitude. Apart from the negative impact through import competition, a more productive China can also benefit the other countries. Similar to the impacts on the OUR, a larger and more productive MUR provides the foreign firms with cheaper intermediate inputs, and foreign consumers with cheaper consumption goods. The presence of both forces in general leaves the welfare impacts on the other countries ambiguous. In the quantification, we find that the welfare impacts on both ROW and ODC are mostly negative but small. However, the impact in welfare for ROW is a quantitative result based on the specific parameterization of the baseline and the counter-factual exercise, rather than a theoretical result from our model. In principle, the model is rich enough to allow for the urbanization policy in China to either reduce or lift the welfare of other countries.

5.1.2 Low λ counter-factual

In the second counter-factual exercise, we extend the relatively liberal migration restrictions on rural-OUR migration to the rural-MUR migration. In practice, this amounts to setting $\lambda_{31} = \lambda_{21} = 5.68$. The resulting migration pattern is reported in Panel (c) of Table 5; the welfare impacts are shown in the third panel of Table 4.

The results are qualitatively similar to the previous exercise with the following notable differences. With lower migration costs to move out, an additional $0.622 - 0.484 = 13.8\%$ of the rural population migrates out as compared to the baseline and λ^* cases. These additional emigrants prefer to move into large cities, and as a result, the population of the large cities increases by $0.6763/0.5209 - 1 = 29.8\%$ relative to the baseline, and $0.6763/0.6209 - 1 = 8.9\%$ relative to the λ^* counter-factual. The further concentration of the workforce in the MUR leads to a more substantial welfare gain in all three regions, and a national welfare gain of $0.2608/0.2256 = 15.6\%$. Relative to the gain in the λ^* counter-factual, the further relaxation of migration restrictions results in a much higher welfare improvement.

In the global context, the negative impacts on foreign firms are intensified compared to the λ^* case: around one third of the operating firms in ODC and ROW are driven out of business. The numbers of exporting firms in the two foreign economies also face a downfall of a similar magnitude. Conversely, these results echo the idea that differential reforms that restrain the growth of large and productive cities may cause firms to move to foreign countries, as emphasized in the introduction.

Second, the welfare impacts outside of China are much richer. In the ODC, both the rural and urban welfare is slightly higher in the counter-factual by a small margin. As mentioned earlier, the positive impact in urban welfare is possible in the model if the benefits from cheaper imports outweigh the loss from import competition.

Lastly, the rural welfare in the ROW *improves* by $7.22/6.88 - 1 = 4.9\%$ whereas the urban welfare *drops* slightly. These changes in welfare are driven mainly by trade in agriculture goods. Under the low λ policy, people move out of the rural region in China, subsequently reducing the agriculture supply; as the migrants earn a higher wage in urban China, the agriculture demand from China surges. These two forces jointly push up the price of food in the international market, and in turn benefit the producers (the rural region) at the expense of the consumers (the urban region) in foreign countries, as we have seen in the ROW. This impact is absent in the ODC because it is not actively trading with China in the agriculture market, as its agriculture productivity is similar to the level of China, as reported in Table 3a.

5.2 Urban Land Rents

As discussed in Section 4.3, other than the migration policy, the changes in urban land rents can also potentially explain the changes in migration patterns between 2005 and 2015. To evaluate the impacts of this channel, we carry out the following counter-factual simulation.

In the baseline quantification, the outside values of land in the MUR grew much faster than the OUR between 2005 and 2015. In 2005, the $\bar{R}_i$ in the MUR (124.24) is only 2.4 times higher than that in the OUR (51.65). However, in 2015, the $\bar{R}_i$ in the MUR (657.16) is already more than five times higher than that in the OUR (125.15). A high land rent that deters immigration could result from the stringent land supply, as well as other factors, in the MUR. To evaluate the impacts of the surging land rents in the MUR, we carry out a "low growth of $\bar{R}_{\text{MUR}}$" counterfactual in which we set the ratios of $\bar{R}_i$ between the MUR and OUR in 2015 to be the same as in the year 2005. This is done by setting $\bar{R}_{\text{MUR}} = 301.04$ while keeping $\bar{R}_{\text{OUR}} = 125.15$, the same as in the baseline. All of the other parameters of the model are the same as in the baseline. The migration patterns of the counter-factual exercise are reported in Table 5, and the welfare impacts in Table 4.

It turns out that the rapid appreciation of land rents can hardly explain the observed migration pattern. As reported in Panel (d) of Table 5, depressing the high land rents in the MUR only increases its inflow from the rural areas by 0.6% to 9.0%. Meanwhile, the rural migrants still predominately prefer the OUR with a 29.1% migration probability.

5.3 Comparing Urbanization Policies with Trade Policies

In Section 5.1, we evaluated the economic impact of the differential reforms on migration friction by simulating two alternative urbanization policies. In this subsection, we compare the urbanization policies to another major category of economic policy: trade policies. Urbanization policies and trade policies target different margins of economic activity: the former concerns the movement of labor while the latter focuses on the mobility of goods. The comparison between the two puts urbanization policies in a familiar perspective based on the vast literature on trade liberalization. Table 6 summarizes all the results, in which we replicate the baseline results in year 2015 in the first panel for ease of exposition.

	China	Rural	MUR	OUR	Rural ODC	Urban ODC	Rural ROW	Urban ROW
	Baseline 2015							
Population	2.1847	0.6629	0.5209	1.0009	1.4761	3.0510	0.0322	1.3668
Welfare	0.2256	0.0609	0.2823	0.2347	0.2747	2.1750	6.8808	10.2857
Operating Firms	2.4687	-	1.2384	1.2303	-	190.8005	-	6231.7030
Exporting Firms	0.4261	-	0.2137	0.2123	-	11.6143	-	54.0165
	Trade Liberalization, λ^*							
Population	2.1847	0.6722	0.5187	0.9938	1.4761	3.0510	0.0322	1.3668
Welfare	0.2314	0.0636	0.2896	0.2408	0.2750	2.1776	6.7820	10.2999
Operating Firms	2.0743	-	1.0420	1.0324	-	159.7666	-	5210.8982
Exporting Firms	0.4149	-	0.2084	0.2065	-	9.7212	-	45.1868
	Trade Liberalization, low λ							
Population	2.1847	0.6980	0.5127	0.9740	1.4761	3.0510	0.0322	1.3668
Welfare	0.2608	0.0750	0.3267	0.2717	0.2765	2.1890	6.4633	10.3540
Operating Firms	1.1092	-	0.5593	0.5499	-	82.8352	-	2688.9079
Exporting Firms	0.3695	-	0.1863	0.1832	-	5.0336	-	23.3476

Table 6: Comparing Urbanization Policies with Trade Policies

Note: This table lists the key endogenous variables for all 7 regions across the baseline and the counter-factual simulations. The first column is the aggregate result for China.

The previous subsection has shown that the λ^* policy, in which the rural-urban migration frictions are equalized while keeping constant the total emigration of the rural population, delivers a 2.6% gain in national welfare. To put this welfare gain into perspective, we solve for a level of trade liberalization that delivers the same 2.6% gain in national welfare. We implement the trade liberalization as uniform reduction in τ_{ij} where either i or j (but not both) is a region in China. In other words, we reduce the variable trade costs in and out of China by a certain percentage, while keeping the trade costs constant everywhere else. All the other parameters, including the migration costs, are kept the same as in the baseline model. The results are presented in the second panel of Table 6. The goal of such an exercise is to find what the λ^* policy amounts to in terms of percentage reduction in bilateral variable trade cost. We carry out a similar exercise to find the level of trade liberalization that delivers a 15.6% improvement in national welfare, the same as in the "low λ" urbanization policy; the results are reported in the third panel of the same table.

The welfare impacts of the urbanization policies are equivalent to substantial reductions in trade friction. To achieve the same level of national welfare improvements in the λ^* counterfactual, China must reduce the bilateral variable trade costs by 5.8%; the equivalent of the "low λ" policy calls for a whopping 24.6% reduction in bilateral iceberg trade costs.[23] These levels of trade liberalizations are notable; by our own calculation based on the ESCAP-World Bank Trade Costs Database, China only lowered its average variable trade costs by 5.1% during 1996–2006, which was when it entered the World Trade Organization (at the end of 2001) and when tariffs were substantially reduced.

The welfare-equivalent trade liberalization is sizable because trade induces only a mild response in internal migration, as evident by comparing the migration matrices reported in Panels (e) and (f) of Table 5 with the data in Table 1. As a large fraction of the population remains in the low-income rural region under trade liberalization, the overall reduction in trade friction must be large enough to achieve the same level of national welfare improvements in the alternative urbanization policies. In comparison, alternative urbanization policies work by diverting the population into high-productivity regions, improving national welfare.

[23]A 5.8% reduction in bilateral trade costs reduces the iceberg trade costs between China and the ODC from 2.46 to $(2.46 - 1) * (1 - 0.058) + 1 = 2.38$. Note that the reduction of τ_{ij} is towards 1, not towards 0, because $\tau_{ij} = 1$ is the limit case of free trade. The other cases are computed in a similar manner.

5.4 Role of International Trade

To understand the role of international trade in the effects of urbanization policy, we re-calibrate the model to the autarky case and re-do the two counter-factual urbanization policies. Comparing these results to the counter-factual simulations under trade offers insight into the interactions between trade and migration. The resulting migration matrices are reported in the last two panels of Table 5 and the welfare results in Table 7.

	China	Rural	MUR	OUR	Rural ODC	Urban ODC	Rural ROW	Urban ROW
	Baseline 2015, Autarky							
Population	2.1847	0.6630	0.5209	1.0008	-	-	-	-
Welfare	0.2005	0.0336	0.2642	0.2185	-	-	-	-
Operating Firms	81.6703	-	41.0766	40.5937	-	-	-	-
Exporting Firms	0.0000	-	0.0000	0.0000	-	-	-	-
	$\lambda_{21} = \lambda_{31} = \lambda^{*}$, Autarky							
Population	2.1847	0.6681	0.6202	0.8964	-	-	-	-
Welfare	0.2060	0.0348	0.2717	0.2217	-	-	-	-
Operating Firms	68.5687	-	39.4698	29.0989	-	-	-	-
Exporting Firms	0.0000	-	0.0000	0.0000	-	-	-	-
	Low λ, Autarky							
Population	2.1847	0.6122	0.6627	0.9098	-	-	-	-
Welfare	0.2279	0.0395	0.2849	0.2321	-	-	-	-
Operating Firms	37.8253	-	22.2478	15.5775	-	-	-	-
Exporting Firms	0.0000	-	0.0000	0.0000	-	-	-	-

Table 7: Impacts of Urbanization Policies under Autarky

Note: This table lists the key endogenous variables for China under the autarky simulation, in which we reduce the model to contain only China. The first column is the aggregate result for China.

International trade increases the overall level of emigration from the rural area. In the low-λ policy in which the entire country adopts a more liberal migration policy, fewer rural migrants choose to move to the urban regions in autarky. In the baseline low-λ simulation, 48.4% of the rural workers choose to stay, and in the autarky-low-λ case, this number increases to 50.7%. The urban regions are less attractive to rural migrants due to the loss of labor demand from exports. The reduced rural-to-urban migration flow also reduces the welfare gain of the low-λ policy reduces to 0.2279/0.2005 - 1 = 13.7%. International trade acting as a magnet that draws rural workers into urban regions is also documented in Fan (2019).

However, the relative flows of rural emigrants to the two urban regions change little. In the autarky-low-λ case, the fraction of rural emigrants to the MUR is 29.1/(100-50.7)=59.0%; this fraction in the baseline low-λ case is 58.7%. By construction, the λ^* counter-factual fixes the total rural emigration, but the same negligible difference in the relative flows between autarky and trade is readily verified. As the spatial distribution of population in the autarky-λ^* case changes little from the baseline-λ^* case, the welfare gains of the λ^* policy remain similar.

5.5 Robustness Checks

In this section, we present four robustness checks. In the first, we use a higher migration elasticity, κ, and in the second one, a higher ε to capture a world with weaker market power. In the third exercise, we experiment with a higher expenditure share of land consumption, and in the last, we shut down the channel of entry and exit of firms. The complete details, including all of the related tables, are relegated to Online Appendix D; the results are briefly described here.

Higher Migration Elasticity In the baseline quantification of the model we use a migration elasticity of $\kappa = 1.63$. Although our choice of κ lies within the range of common estimates between 1.4 and 3.3 in the literature, it nevertheless is closer to the lower end. As a robustness check, we re-calibrate $\{\lambda_{ij}\}$ and f_e in the year 2015 using $\kappa = 3.3$ from Monte et al. (2018), the estimate on the higher end.

A higher migration elasticity implies that the estimated λ_{ij} are smaller in levels and less dispersed. The key pattern is still preserved in the case with higher κ: in 2015, it is significantly harder to move from the rural regions to the large cities than to the smaller ones. The impacts of the alternative urbanization policies are qualitatively similar but quantitatively larger. Adopting the λ^* policy leads to 3.0%, and the "low λ" policy, a 18.2% increase in national welfare. These numbers are to be compared with the 2.6% and 15.6% welfare gains in the baseline. The welfare gains are higher here because the migration flows are more sensitive to the changes in λ_{ij} in a world with a high elasticity.

Higher Elasticity of Substitution In the baseline model, we jointly calibrate ε and θ to match a trade elasticity of 4 and a tail-index of firm-size distribution of 1.076. The resulting $\varepsilon = 4.717$ implies an average markup of 27%. In the robustness check, we increase the elasticity of substitution to $\varepsilon = 10.0$ so the market structure is closer to perfect competition with a markup of 11% while the trade elasticity, θ, is reset to 9.684 to match the tail-index of 1.076. The new values of ε and θ remain in the ballpark of the estimates from the gravity-equation literature.

With a lower markup, the real income level in all cases improves substantially, as lower market power increases the firms' equilibrium output. In this framework in which there is a differentiated sector with positive markups and a rural sector with zero markups, the equilibrium allocation is always sub-optimal as the allocation of labor to urban regions is less than optimal. A reduction in market power in the differentiated sector reduces this allocative inefficiency and implies a larger rural-urban migration.[24] However, to keep migration flows as the observed ones, the re-calibrated rural-to-urban migration frictions would be higher than the baseline ones. Similarly, the urban-to-rural migration frictions are lower than the baseline ones.

The λ^* and the low-λ counterfactuals lead to the slightly lower welfare gains of 1.8% and 10.9%. In this framework, the elasticity of welfare to the allocation of labor is tied closely to the elasticity of substitution, which inversely reflects the love of variety. In the λ^* counter-factual, labor reallocation from the OUR to the MUR still brings welfare gains, but such gains become smaller when ϵ is higher because the new varieties that come with the inflow of population to the MUR are less valuable to consumers there. A similar logic applies to the low-λ counter-factual.

Higher Expenditure Share of Land Consumption In the baseline model, the expenditure share of land consumption, $1 - \alpha - \gamma$, is set to 0.0625 following Davis and Ortalo-Magné (2011) and Combes et al. (2019), as explained in Section 4.2.1. As this paper does not explicitly model housing structure, which is treated as part of the differentiated goods, we consider this number an appropriate one to use. The finding that households are not entirely responsive to land prices may be likely because this expenditure share is low. Thus, we conduct a robustness check by experimenting with a higher expenditure share at 0.25, which can be considered on the high end of possible values of the expenditure share of land consumption.

With a higher weight on land consumption, the MUR and OUR's high land prices deter rural migrants. As a result, the estimated rural-to-urban migration frictions drastically decline. Never-

[24]For the economics underlying allocative inefficiency due to variable markups, see, for example, Holmes et al. (2014) and Arkolakis et al. (2019).

theless, the migration barriers into the MUR are still higher than that into the OUR, reflecting the discriminatory urbanization policy. We first repeat the "low growth of $\bar{R}_{\text{MUR}}$" counterfactual exercise. Naturally, we find that rural emigrants are more responsive to the changes in land prices than the baseline model. With a lower growth rate of $\bar{R}_{\text{MUR}}$, the rural-to-MUR migration probability increased to 11.1%. In comparison, in the baseline model, the same probability is only 9.0%.

Our main results are robust to this alternative parameterization. The λ^* and low-λ policies still divert a significant proportion of the rural emigrants towards the MUR, but the magnitudes are smaller than the baseline case. Importantly, the alternative urbanization policies are still more effective in re-directing the population flows than depressing the growth rates of $\bar{R}_{\text{MUR}}$. The welfare impacts of the alternative urbanization policies become relatively mild. This is because the alternative policies attract a smaller fraction of the rural emigrants into the more productive MUR if the individuals care more about land consumption.

Fixed Entry The firm-entry margin is instrumental to the punchline result that a more uniform or laissez-faire migration policy improves national welfare. To highlight the role of firm entry in our model, we shut down the firms' entry-and-exit channel. In the baseline model, I_j potential firms pay the entry fee f_e, and we need to solve for I_j in the general equilibrium. For the "fixed entry" model, we assume that $\bar{I}_j$ is exogenously given at a level that will be specified later, and the entry cost f_e is assumed to be zero. Without the firm-entry margin, the aggregate profit becomes positive (instead of zero in the baseline model). Regional aggregate profits are evenly rebated to individuals in that region. The details of solving the model in this new setup are provided in Online Appendix B.7.

To study the effect of entry on migration, we compute $\bar{I}_j$'s used in all fixed-entry exercises from the pre-migration equilibrium, i.e., the equilibrium under the initial population and before people move. In practice, this is computed from the baseline model with free entry and prohibitive migration costs ($\lambda_{ij} = \infty, i \neq j$). Then, any equilibrium with the fixed $\bar{I}_j$'s is a world in which the number of entrants no longer responds to population flows.

The estimated λ_{ij}'s are similar to the baseline case; the λ^* exercise also leads to a pronounced shift of the rural emigrants towards the MUR, although the magnitude is slightly smaller. Nevertheless, the welfare impacts of alternative policies are drastically different. In the fixed-entry model, the MUR suffers lower welfare when more rural migrants flow into the large cities in the λ^* exercise. As a result, national welfare drops. This is in stark contrast with the result in the baseline model in which all regional and national welfare increases. As this result in the baseline model is the paper's punchline, this exercise under the fixed-entry model highlights the importance of the entry margin. Under the fixed mass of firms, an increase in the population of a region due to migration no longer increases the number of varieties there; instead, they only push down the wages, push up the land prices, and eventually reduce local welfare. The importance of firm entry in the context of migration is already highlighted in Ma and Tang (2020b), and this robustness check resonates with their finding. The result in the low-λ policy is similar in terms of the directions of changes in regional welfare, but the national welfare still improves as the overall migration frictions are lowered.

6 Conclusion

This paper documents a striking contrast in the migration patterns between years 2005 and 2015: whereas the migration probability of a rural individual to mega cities is higher than that to smaller cities in 2005, the opposite is true in 2015. Such a reversal in the migration pattern is consistent with the differential reforms to the hukou system that encourage rural people to move into small- and medium-sized cities and restrain the growth of large cities.

In our quantitative analysis, we find that equalizing migration costs between rural-MUR and rural-OUR migration while keeping the total emigration of the rural population unchanged results in welfare gains of 2.6%. This is substantial, considering that it involves reallocating a mere 4.4% of the total population. We also find that a more laissez-faire urbanization policy that equalizes the rural-MUR migration cost to the rural-OUR one results in welfare gains of 15.6%. The welfare gains under the two alternative urbanization policies amount to what would result from 5.8% and 24.6% reductions in bilateral variable trade costs, respectively. Based on the ESCAP-World Bank Trade Costs Database, China only lowered its average iceberg trade costs by 5.1% during 1996–2006, which was when it entered the World Trade Organization (at the end of 2001) and tariffs were substantially reduced. Namely, an alternative urbanization policy that treats large and small cities equally while keeping the total rural emigration the same increases welfare by a similar magnitude to the trade liberalization that it has accomplished in the past. China can gain even more with a more laissez-faire urbanization policy.

Recall that our welfare results are conservative estimates because of the relatively low migration elasticity adopted in the baseline. As shown in Section 5.5, the welfare gains increase substantially if the migration elasticity is higher.

Our quantitative analyses are, of course, specific to our model. However, our model is mostly standard. For the sake of tractability, the model does not incorporate agglomeration forces in cities despite the idiosyncratic locational preferences and the urban costs being dispersion forces. It would be an interesting extension to incorporate agglomeration forces; if the net effect is positive, i.e., some sorts of increasing returns at the city level exist, the message of this paper would become even stronger because the large, productive urban centers in the MUR would become all the more important in enhancing aggregate welfare.

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.

An, Lei, Yu Qin, Jing Wu, and Wei You, "The Local Labor Market Effect of Relaxing Internal Migration Restrictions: Evidence from China," *Available at SSRN 3626292*, 2020.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare, "The elusive pro-competitive effects of trade," *The Review of Economic Studies*, 2019, *86* (1), 46–80.

Artuç, Erhan, Shubham Chaudhuri, and John McLaren, "Trade Shocks and Labor Adjustment: A Structural Empirical Approach," *The American Economic Review*, 2010, *100* (3), 1008–1045.

Axtell, Robert L, "Zipf Distribution of U.S. Firm Sizes," *Science*, 2001, *293* (5536), 1818–1820.

Bryan, Gharad and Melanie Morten, "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268.

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, *87* (3), 741–835.

Chaney, Thomas, "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, September 2008, *98* (4), 1707–21.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, "The costs of agglomeration: House and land prices in French cities," *The Review of Economic Studies*, 2019, *86* (4), 1556–1589.

Davis, Morris A and François Ortalo-Magné, "Household expenditures, wages, rents," *Review of Economic Dynamics*, 2011, *14* (2), 248–261.

di Giovanni, Julian and Andrei A. Levchenko, "Country Size, International Trade, and Aggregate Fluctuations in Granular Economies," *Journal of Political Economy*, 2012, *120* (6), 1083–1132.

_ , _ , and Francesc Ortega, "A Global View Of Cross-Border Migration," *Journal of the European Economic Association*, 02 2015, *13* (1), 168–202.

Donaldson, Dave and Richard Hornbeck, "Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 02 2016, *131* (2), 799–858.

Du, Jinfeng and Richard B Peiser, "Land supply, pricing and local governments' land hoarding in China," *Regional Science and Urban Economics*, 2014, *48*, 180–189.

Fajgelbaum, Pablo D, Eduardo Morales, Juan Carlos Suárez Serrato, and Owen Zidar, "State Taxes and Spatial Misallocation," *The Review of Economic Studies*, 2018, *86* (1), 333–376.

Fan, Jingting, "Internal Geography, Labor Mobility, and the Distributional Impacts of Trade," *American Economic Journal: Macroeconomics*, July 2019, *11* (3), 252–88.

Fujita, Masahisa, Tomoya Mori, J Vernon Henderson, and Yoshitsugu Kanemoto, "Spatial distribution of economic activities in Japan and China," in "Handbook of regional and urban economics," Vol. 4, Elsevier, 2004, pp. 2911–2977.

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

Harris, John R. and Michael P. Todaro, "Migration, Unemployment and Development: A Two-Sector Analysis," *The American Economic Review*, 1970, *60* (1), 126–142.

Holmes, Thomas J, Wen-Tai Hsu, and Sanghoon Lee, "Allocative efficiency, mark-ups, and the welfare gains from trade," *Journal of International Economics*, 2014, *94* (2), 195–206.

Hsieh, Chang-Tai and Enrico Moretti, "Housing Constraints and Spatial Misallocation," *American Economic Journal: Macroeconomics*, April 2019, *11* (2), 1–39.

_ and Peter J Klenow, "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

Hsu, Wen-Tai, Xiaolu Li, Yang Tang, and Jing Wu, "Determinants of urban land supply in the people's republic of China: how do political factors matter?," *Asian Development Review*, 2017, *34* (2), 152–183.

Lewis, Arthur, "Economic Development with Unlimited Supplies of Labour," *The Manchester School*, 1954, *22* (2), 139–191.

Luttmer, Erzo G. J., "Selection, Growth, and the Size Distribution of Firms," *The Quarterly Journal of Economics*, 08 2007, *122* (3), 1103–1144.

Ma, Lin and Yang Tang, "The Distributional Impacts of Transportation Networks in China," Technical Report, working paper 2020.

_ **and** _ , "Geography, trade, and internal migration in China," *Journal of Urban Economics*, 2020, *115*, 103181. Cities in China.

Melitz, M.J., "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 2003, *71* (6), 1695–1725.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg, "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, December 2018, *108* (12), 3855–90.

Murata, Yasusada, "Product diversity, taste heterogeneity, and geographic distribution of economic activities:: market vs. non-market interactions," *Journal of Urban Economics*, 2003, *53* (1), 126 – 144.

Novy, Dennis, "Gravity Redux: Measuring International Trade Costs with Panel Data," *Economic Inquiry*, 2013, *51* (1), 101–121.

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

Redding, Stephen J., "Goods trade, factor mobility and welfare," *Journal of International Economics*, 2016, *101*, 148 – 167.

_ **and Esteban Rossi-Hansberg**, "Quantitative spatial economics," *Annual Review of Economics*, 2017, *9*, 21–58.

Simonovska, Ina and Michael E. Waugh, "The elasticity of trade: Estimates and evidence," *Journal of International Economics*, 2014, *92* (1), 34–50.

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti, "Growing Like China," *American Economic Review*, February 2011, *101* (1), 196–233.

Stillwell, John, Konstantinos Daras, Martin Bell, and Nik Lomax, "The IMAGE studio: A tool for internal migration analysis and modelling," *Applied Spatial Analysis and Policy*, 2014, *7* (1), 5–23.

Tombe, Trevor and Xiaodong Zhu, "Trade, Migration, and Productivity: A Quantitative Analysis of China," *American Economic Review*, May 2019, *109* (5), 1843–72.

Zi, Yuan, "Trade liberalization and the great labor reallocation," Technical Report 2020. CEPR working paper DP14490.

Online Appendix for "Urbanization Policy and Economic Development: A Quantitative Analysis of China's Differential Hukou Reforms"

Wen-Tai Hsu* Lin Ma†

December 2020

A Supplementary Tables

Beijing	Tianjin	Shijiazhuang	Shenyang
Haerbin	Shanghai	Nanjing	Suzhou
Hangzhou	Ningbo	Wenzhou	Qingdao
Zhengzhou	Wuhan	Guangzhou	Shenzhen
Foshan	Dongguan	Chongqing	Chengdu
Xi'an			

Table A.1: Cities in the MUR

Note: This table lists all the prefecture-level cities that are included in the MUR. These are the cities with an urban population greater than 5 million in the 2010 census. All the other prefecture cities are grouped into the OUR.

*School of Economics, Singapore Management University. Email: wentaihsu@smu.edu.sg.
†Department of Economics, National University of Singapore. Email: ecsml@nus.edu.sg.

ARG	CHE*	DNK*	HRV	JPN*	MEX	POL	SWE*
AUS*	CHL	ESP	HUN	KAZ	MLT	PRT	THA
AUT*	CHN	EST	IDN	KHM	MYS	ROU	TUN
BEL*	COL	FIN*	IND	KOR	NLD*	RUS	TUR
BGR	CRI	FRA*	IRL*	LTU	NOR*	SAU*	USA*
BRA	CYP	GBR*	ISL*	LUX*	NZL	SGP*	VNM
BRN*	CZE	GRC	ISR	LVA	PER	SVK	ZAF
CAN*	DEU*	HKG*	ITA*	MAR	PHL	SVN	

Table A.2: Country List

Note: This table lists all the countries in the quantitative exercise. The countries with a star are included in the ROW; those without, except for China, are included in the ODC. The reported codes are the ISO 3166 alpha-3 country codes. More details can be found at `https://www.iso.org/iso-3166-country-codes.html`.

B Model Solution

The equilibrium conditions in the model can be described as a system of nonlinear equations in which $\{w_j, I_j, P_j, N_j\}$ are the endogenous variables to be solved. We solve the system of equations with iterations: in the current iteration, the system of equations implies new values of $\{w_j, I_j, P_j, N_j\}$ as functions of the current values. The algorithm continues until the current and implied values of endogenous variables converge under a pre-specified tolerance level, 1.0E-6. In this appendix, we describe the equations and rules to update each variable above. Before venturing into each variable in detail, we define notation and highlight conditions that will be used across the entire algorithm.

Notations

1. In describing the iterative method, we denote the values in the current iteration as x, and the implied values as x'.

2. We define the set of the urban locations as $\mathcal{U}$ and the rural locations as $\mathcal{R}$ with the understanding that $\mathcal{U} \cup \mathcal{R}$ covers all the locations, and $\mathcal{U} \cap \mathcal{R} = \emptyset$.

3. For computational reasons, we use the Υ matrix to denote a combination of trade costs. The element in the i-th row and j-th column is

$$\Upsilon_{ij} = (\tau_{ij})^{-\theta} (f_{ij})^{-\frac{\theta-\varepsilon+1}{\varepsilon-1}} .$$

4. As $\frac{1}{a}$ follows a Pareto distribution, we are working with the following CDF and PDF of a:

$$G_j(a) = \mu_j^\theta a^\theta$$
$$g_j(a) = \theta \mu_j^\theta a^{\theta-1}.$$

Income The free-entry condition implies that the total profit in each urban region is zero. As a result, the total income in region j is the labor income inclusive of land rents, $(w_j + T_c)N_j$, where c is the country to which j belongs. The total income in the rural regions adopts the same expression due to the perfectly competitive agriculture market.

Land Rents Cost minimization in the agriculture sector implies that:

$$R_{A,c}L_{A,c} = \frac{\eta}{\nu} w_j N_j,$$

where j is the rural region in country c. The aggregate land rent is then computed as

$$\begin{aligned} T_c &= \frac{R_{A,c}L_{A,c} + (1-\alpha-\gamma)\sum_{i\in J_c} w_i N_i}{(\alpha+\gamma)\bar{N}_c} \\ &= \frac{\eta\nu^{-1} w_j N_j + (1-\alpha-\gamma)\sum_{i\in J_c} w_i N_i}{(\alpha+\gamma)\bar{N}_c}. \end{aligned}$$

Equivalently, we can also express the aggregate land rent as the sum of the rent from the rural and the urban areas:

$$T_c \bar{N}_c = \frac{\left(\frac{\eta}{\nu} + 1 - \alpha - \gamma\right) w_j N_j + (1-\alpha-\gamma)\sum_{i\in J_c \cap \mathcal{U}} w_i N_i}{\alpha+\gamma} \tag{B.1}$$

Expenditure Out of the total income, a fraction γ is spent on differentiated goods by consumers. Moreover, firms also demand differentiated products as inputs in both urban and rural regions. As a result, the total expenditure on differentiated products, X_j, comes from both parts in urban locations. In the urban regions, the expenditure can be expressed as

$$X_j = \gamma(w_j + T_c)N_j + (1-\beta)X_j = \frac{\gamma}{\beta}(w_j + T_c)N_j, \quad j \in \mathcal{U}. \tag{B.2}$$

The inputs to produce X_j worth of differentiated products equal $(1-\beta)X_j$. This observation relies on the fact that the total profit in the differentiated sector equals zero, so that the total revenue equals the total costs in region j.

In rural regions, X_j depends on consumer demand and the demand from the agriculture sector:

$$X_j = \gamma(w_j + T_c)N_j + \frac{1-\nu-\eta}{\nu} w_j N_j. \tag{B.3}$$

The second term in the expression above captures the demand from the agriculture sector. Note that the total input costs of the agriculture sector must be $w_j N_j/\nu$ in equilibrium, and a fraction $1-\nu-\eta$ of the costs is used to purchase intermediate inputs.

The expenditure on the agriculture products is $X_j^A = \alpha(w_j + T_c)N_j$ in each location j.

The total expenditure of the country c, X_c, is the summation of the expenditures of all the regions: $X_c = \sum_{j\in J_c} X_j$, where J_c is the set of regions in country c.

B.1 Updating P_j

We can explicitly write the ideal price index in the differentiated sector as

$$\begin{aligned}
P_j &= \left[\sum_{i\in\mathcal{U}} \left(\frac{\varepsilon}{\varepsilon-1}\tau_{ji}\chi_i\right)^{1-\varepsilon} I_i \int_0^{a_{ji}} a^{1-\varepsilon} g_i(a)da\right]^{\frac{1}{1-\varepsilon}} \\
&= \left[\sum_{i\in\mathcal{U}} \left(\frac{\varepsilon}{\varepsilon-1}\tau_{ji}\chi_i\right)^{1-\varepsilon} I_i \frac{\theta}{\theta-(\varepsilon-1)} (\mu_i)^{\theta} (a_{ji})^{\theta-(\varepsilon-1)}\right]^{\frac{1}{1-\varepsilon}} \\
&= \left[\sum_{i\in\mathcal{U}} \left(\frac{\varepsilon}{\varepsilon-1}\tau_{ji}\chi_i\right)^{1-\varepsilon} I_i \frac{\theta}{\theta-(\varepsilon-1)} (\mu_i)^{\theta} \left(\frac{\varepsilon-1}{\varepsilon}\frac{P_j}{\tau_{ji}\chi_i}\left(\frac{X_j}{\varepsilon\chi_i f_{ji}}\right)^{\frac{1}{\varepsilon-1}}\right)^{\theta-(\varepsilon-1)}\right]^{\frac{1}{1-\varepsilon}}
\end{aligned}$$

$$(P_j)^{\frac{\theta}{\varepsilon-1}} = \left(\frac{\varepsilon}{\varepsilon-1}\right)^{\frac{\theta}{\varepsilon-1}} \left(\frac{\theta}{\theta-(\varepsilon-1)}\right)^{\frac{1}{1-\varepsilon}} \left(\frac{X_j}{\varepsilon}\right)^{\frac{\theta-(\varepsilon-1)}{(\varepsilon-1)(1-\varepsilon)}} \left[\sum_{i\in\mathcal{U}} I_i \left(\frac{\mu_i}{\tau_{ji}\chi_i}\right)^{\theta} \left(\frac{1}{\chi_i f_{ji}}\right)^{\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}\right]^{\frac{1}{1-\varepsilon}};$$

therefore the rule to update P_j, conditional on X_j, I_i, and χ_i, is

$$P_j' = \frac{\varepsilon}{\varepsilon-1}\left(\frac{\theta}{\theta-(\varepsilon-1)}\right)^{-\frac{1}{\theta}} \left(\frac{X_j}{\varepsilon}\right)^{-\frac{\theta-(\varepsilon-1)}{\theta(\varepsilon-1)}} \left[\sum_{i\in\mathcal{U}} I_i\left(\Upsilon_{ji}\right)(\mu_i)^{\theta}(\chi_i)^{-\theta-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}\right]^{-\frac{1}{\theta}}. \tag{B.4}$$

Note that due to the assumption of free internal trade, the price level only varies at the country level. As a result, we can also express the price as

$$P_c' = \frac{\varepsilon}{\varepsilon - 1} \left(\frac{\theta}{\theta - (\varepsilon - 1)} \right)^{-\frac{1}{\theta}} \left(\frac{X_c}{\varepsilon} \right)^{-\frac{\theta - (\varepsilon - 1)}{\theta(\varepsilon - 1)}} \left[\sum_{i \in \mathcal{U}} I_i \left(\Upsilon_{ci} \right) \left(\mu_i \right)^{\theta} \left(\chi_i \right)^{-\theta - \frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}} \right]^{-\frac{1}{\theta}}. \tag{B.5}$$

B.2 Trade Flow

Denote the sales of the differentiated products from j to i as X_{ij}. We can express X_{ij} as

$$\begin{aligned}
X_{ij} &= I_j \int_0^{a_{ij}} p_{ij}(a) q_{ij}(a) dG_j(a) \\
&= I_j \int_0^{a_{ij}} \frac{X_i}{(P_i)^{1-\varepsilon}} \left[p_{ij}(k) \right]^{1-\varepsilon} dG_j(a) \\
&= I_j \frac{X_i}{(P_i)^{1-\varepsilon}} \left[\frac{\varepsilon}{\varepsilon - 1} \tau_{ij} \chi_j \right]^{1-\varepsilon} \int_0^{a_{ij}} a^{1-\varepsilon} dG_j(a) \\
&= I_j \frac{X_i}{(P_i)^{1-\varepsilon}} \left[\frac{\varepsilon}{\varepsilon - 1} \tau_{ij} \chi_j \right]^{1-\varepsilon} \frac{\theta}{\theta - (\varepsilon - 1)} \left(\mu_j \right)^{\theta} \left(a_{ij} \right)^{\theta - (\varepsilon - 1)} \\
&= I_j \frac{X_i}{(P_i)^{1-\varepsilon}} \left[\frac{\varepsilon}{\varepsilon - 1} \tau_{ij} \chi_j \right]^{1-\varepsilon} \frac{\theta}{\theta - (\varepsilon - 1)} \left(\mu_j \right)^{\theta} \left[\frac{\varepsilon - 1}{\varepsilon} \frac{P_i}{\tau_{ij} \chi_j} \left(\frac{X_i}{\varepsilon \chi_j f_{ij}} \right)^{\frac{1}{\varepsilon - 1}} \right]^{\theta - (\varepsilon - 1)} \\
&= I_j \left[\frac{X_i}{(P_i)^{1-\varepsilon}} \right]^{\frac{\theta}{\varepsilon - 1}} \left(\frac{\varepsilon - 1}{\varepsilon} \right)^{\theta} \left(\tau_{ij} \right)^{-\theta} \left(\chi_j \right)^{-\theta - \frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}} \left(f_{ij} \right)^{-\frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}} \frac{\theta}{\theta - (\varepsilon - 1)} \left(\mu_j \right)^{\theta} \varepsilon^{-\frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}} \\
&= I_j \left(X_i \right)^{\frac{\theta}{\varepsilon - 1}} \left(P_i \right)^{\theta} \left(\frac{\varepsilon - 1}{\varepsilon} \right)^{\theta} \Upsilon_{ij} \left(\chi_j \right)^{-\theta - \frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}} \frac{\theta}{\theta - (\varepsilon - 1)} \left(\mu_j \right)^{\theta} \varepsilon^{-\frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}}.
\end{aligned}$$

Substitute in the expression of P_i from equation (B.4):

$$\begin{aligned}
X_{ij} &= \frac{I_j \left(X_i \right)^{\frac{\theta}{\varepsilon - 1}} \left(\frac{\varepsilon - 1}{\varepsilon} \right)^{\theta} \Upsilon_{ij} \left(\chi_j \right)^{-\theta - \frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}} \frac{\theta}{\theta - (\varepsilon - 1)} \left(\mu_j \right)^{\theta} \varepsilon^{-\frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}}}{\left[\frac{\varepsilon}{\varepsilon - 1} \left(\frac{\theta}{\theta - (\varepsilon - 1)} \right)^{-\frac{1}{\theta}} \left(\frac{X_i}{\varepsilon} \right)^{-\frac{\theta - (\varepsilon - 1)}{\theta(\varepsilon - 1)}} \left[\sum_{k \in \mathcal{U}} I_k \left(\Upsilon_{ik} \right) \left(\mu_k \right)^{\theta} \left(\chi_k \right)^{-\theta - \frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}} \right]^{-\frac{1}{\theta}} \right]^{-\theta}} \\
&= \frac{I_j \Upsilon_{ij} \left(\mu_j \right)^{\theta} \left(\chi_j \right)^{-\theta - \frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}}}{\sum_{k \in \mathcal{U}} I_k \Upsilon_{ik} \left(\mu_k \right)^{\theta} \left(\chi_k \right)^{-\theta - \frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}}} X_i.
\end{aligned} \tag{B.6}$$

B.3 Updating w_j in the Urban Regions

The total income in an urban region j is the sum from two parts. The first part is the value-added from the differentiated sector, which is the sales to all the urban and the rural regions, minus the costs of the intermediate products. The second part is the residential land rent, which is a fraction of the total income:

$$(w_j + T_c)N_j = \sum_{i\in\mathcal{R}} X_{ij} + \sum_{i\in\mathcal{U}} X_{ij} - (1-\beta)\, X_j + (1-\alpha-\gamma)(w_j + T_c)N_j.$$

In the above equation, the LHS is the total income. The trade balance condition between the rural and urban regions also implies that the total sales to the rural regions must be the same as the total imports of food, and therefore $\sum_{i\in\mathcal{R}} X_{ij} = \alpha(w_j + T_c)N_j$. Substitute this into the equation above:

$$\gamma(w_j + T_c)N_j + (1-\beta)\, X_j = \sum_{i\in\mathcal{U}} X_{ij}.$$

Substitute in the expression of X_j from equation (B.2):

$$\begin{aligned}
\gamma(w_j + T_c)N_j + (1-\beta)\, \frac{\gamma}{\beta}(w_j + T_c)N_j &= \sum_{i\in\mathcal{U}} X_{ij} \\
\frac{\gamma}{\beta}(w_j + T_c)N_j &= \sum_{i\in\mathcal{U}} X_{ij}.
\end{aligned}$$

In the end, substitute in the solution of urban-to-urban trade flows from equation (B.6):

$$\begin{aligned}
\frac{\gamma}{\beta}(w_j + T_{c(j)})N_j &= \sum_{i\in\mathcal{U}} \frac{I_j\left(\mu_j\right)^{\theta}\tau_{ij}^{-\theta}\left(f_{ij}\right)^{-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}\left[w_j^{\beta}\left(P_j\right)^{1-\beta}\right]^{-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}}{\sum_{k\in\mathcal{U}} I_k\left(\mu_k\right)^{\theta}\tau_{ik}^{-\theta}\left(f_{ik}\right)^{-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}\left[w_k^{\beta}\left(P_k\right)^{1-\beta}\right]^{-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}}\frac{\gamma}{\beta}(w_i + T_{c(i)})N_i \\
(w_j + T_{c(j)})N_j &= \sum_{i\in\mathcal{U}} \frac{I_j\Upsilon_{ij}\left(\mu_j\right)^{\theta}\left(\chi_j\right)^{-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}}{\sum_{k\in\mathcal{U}} I_k\Upsilon_{ik}\left(\mu_k\right)^{\theta}\left(\chi_k\right)^{-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}}(w_i + T_{c(i)})N_i. \qquad \text{(B.7)}
\end{aligned}$$

B.4 Updating w_j in the Rural Regions

The wage rates in the rural areas, on the other hand, are determined through the market clearing condition in the agriculture market. We first note that given a rural wage, w_j, and a price index of the differentiated inputs P_j, the costs of input bundle in j becomes:

$$\begin{aligned}\chi_j^A &= (w_j)^\nu (R_{A,c})^\eta (P_j)^{1-\nu-\eta} \\ &= (w_j)^\nu \left(\frac{\eta}{\nu}\frac{N_j}{L_j}w_j\right)^\eta (P_j)^{1-\nu-\eta} \\ &= \left(\frac{\eta}{\nu}\frac{N_j}{L_j}\right)^\eta (w_j)^{\nu+\eta} (P_j)^{1-\nu-\eta}. \end{aligned} \tag{B.8}$$

The wage rate in rural China is the numeraire in our model, and therefore we must solve for the two other rural wage rates to clear the market. The market clearing condition is characterized by the following two equations:

1. If country c does not engage in the international trade in the agricultural products, e.g, all the rural and the urban regions in country c buy agricultural products only from their own rural region, and its rural region sells only domestically, then the rural wage rate, w_j, is determined by the market clearing condition:

$$\begin{aligned} w_j N_j &= \nu \left[\alpha \sum_{i\in J_c} (w_i + T_c) N_i\right] \\ &= \alpha\nu(w_j + T_c)N_j + \nu \left[\alpha \sum_{i\in J_c \cap \mathcal{U}} (w_i + T_c) N_i\right] \\ (1-\alpha\nu)w_j N_j &= \alpha\nu T_c N_j + \nu \left[\alpha \sum_{i\in J_c \cap \mathcal{U}} (w_i + T_c) N_i\right]. \end{aligned}$$

   In this equation, $w_i N_i$ is the total labor income of the rural region, set J_c is the set of the regions that belongs to country c. The terms in the square bracket on the left-hand side (RHS) of the equation is the total expenditure on agricultural goods of all the regions in country c, and ν captures the share of the expenditure that goes to the rural workers.

   In the expression above, T_c is also a function of the rural wage, w_j. Rearrange the equation and substitute in the expression of T_c from equation (B.1), we can express the rural wage rate as a function of the urban wage rates:

$$\begin{aligned}
\frac{(1-\alpha\nu)}{\alpha\nu} w_j N_j &= T_c N_j + \left[\sum_{i\in J_c\cap\mathcal{U}} (w_i + T_c) N_i\right] \\
&= T_c \bar{N}_c + \left[\sum_{i\in J_c\cap\mathcal{U}} w_i N_i\right] \\
&= \frac{\left(\frac{\eta}{\nu} + 1 - \alpha - \gamma\right) w_j N_j + (1-\alpha-\gamma)\sum_{i\in J_c\cap\mathcal{U}} w_i N_i}{(\alpha+\gamma)} + \sum_{i\in J_c\cap\mathcal{U}} w_i N_i.
\end{aligned}$$

Simplify:

$$\begin{aligned}
\left[\frac{(1-\alpha\nu)}{\alpha\nu} - \frac{\frac{\eta}{\nu}+1-\alpha-\gamma}{\alpha+\gamma}\right] w_j N_j &= \frac{1}{\alpha+\gamma}\sum_{i\in J_c\cap\mathcal{U}} w_i N_i \\
w_j N_j &= \frac{\frac{1}{\alpha+\gamma}\sum_{i\in J_c\cap\mathcal{U}} w_i N_i}{\frac{(1-\alpha\nu)}{\alpha\nu} - \frac{\frac{\eta}{\nu}+1-\alpha-\gamma}{\alpha+\gamma}} \\
&= \frac{\sum_{i\in J_c\cap\mathcal{U}} w_i N_i}{(\alpha+\gamma)\frac{1-\alpha\nu}{\alpha\nu} - \left(\frac{\eta}{\nu}+1-\alpha-\gamma\right)} \\
&= \frac{\sum_{i\in J_c\cap\mathcal{U}} w_i N_i}{(\alpha+\gamma)\frac{1}{\alpha\nu} - \left(\frac{\eta}{\nu}+1\right)}.
\end{aligned}$$

2. If country c imports agricultural products from country d, then the agricultural input costs between the two countries must satisfy this equation:

$$\frac{\chi_c^A}{\mu_c^A} = \frac{\tau_{cd}^A \chi_d^A}{\mu_d^A}. \tag{B.9}$$

The LHS is the price of domestic agricultural products in country c, and the RHS is the price of the imported products from country d. We cannot have $\frac{\chi_c^A}{\mu_c^A} < \frac{\tau_{cd}^A \chi_d^A}{\mu_d^A}$ as it would imply that country c should not import from country d. We cannot have $\frac{\chi_c^A}{\mu_c^A} > \frac{\tau_{cd}^A \chi_d^A}{\mu_d^A}$ either, as this implies that the rural region in country c cannot offer a competitive price in its own market despite the trade barrier. If the inequality were true, we could then infer that all the regions in the world would find the price from country d to be lower than the price from country c, and thus the demand for the agriculture goods in country c would drop to zero. This cannot happen in equilibrium because there will always be a strictly positive supply of agricultural products due to the existence of idiosyncratic location preferences.

The above conditions fully characterize the solution to the market clearing conditions in the agricultural market, conditional on a given set of trade relationships (e.g., who imports from whom). In practice, given our 3-country setup, as there is a small number of possible trade relationships, we use a guess-and-verify method to find the equilibrium trade relationships and the corresponding wage rates in the rural regions.

B.5 Updating I_j

The free entry condition in equation (11) in the urban area comes down to

$$\sum_{i=1}^{J}\left\{\frac{X_i}{\varepsilon\left(P_i\right)^{1-\varepsilon}}\left(\frac{\varepsilon}{\varepsilon-1}\tau_{ij}\chi_j\right)^{1-\varepsilon}\frac{\theta\mu_j^{\theta}\left(a_{ij}\right)^{1+\theta-\varepsilon}}{\theta-(\varepsilon-1)}-\mu_j^{\theta}\left(a_{ij}\right)^{\theta}\chi_j f_{ij}\right\}=\chi_j f_e,$$

where the left-hand side is the expected profit, and a_{ij} is the cut-off productivity:

$$\begin{aligned} a_{ij} &= \frac{\varepsilon-1}{\varepsilon}\frac{P_i}{\tau_{ij}\chi_j}\left(\frac{X_i}{\varepsilon\chi_j f_{ij}}\right)^{\frac{1}{\varepsilon-1}} \\ &= \frac{\varepsilon-1}{\varepsilon}\left(\varepsilon\right)^{\frac{1}{1-\varepsilon}}P_i\left(X_i\right)^{\frac{1}{\varepsilon-1}}\left(\chi_j\right)^{\frac{1}{1-\varepsilon}}\left(\frac{\Upsilon_{ij}}{f_{ij}}\right)^{\frac{1}{\theta}}. \end{aligned}$$

Substitute the expression of a_{ij} into the zero-profit condition, and simplify:

$$\begin{aligned} \chi_j f_e &= \sum_{i=1}^{J}\frac{X_i}{\varepsilon\left(P_i\right)^{1-\varepsilon}}\left(\frac{\varepsilon}{\varepsilon-1}\tau_{ij}\chi_j\right)^{1-\varepsilon}\frac{\theta\mu_j^{\theta}\left(\frac{\varepsilon-1}{\varepsilon}\frac{P_i}{\tau_{ij}\chi_j}\left(\frac{X_i}{\varepsilon\chi_j f_{ij}}\right)^{\frac{1}{\varepsilon-1}}\right)^{1+\theta-\varepsilon}}{\theta-(\varepsilon-1)} \\ &\quad -\sum_{i=1}^{J}\mu_j^{\theta}\left(\frac{\varepsilon-1}{\varepsilon}\frac{P_i}{\tau_{ij}\chi_j}\left(\frac{X_i}{\varepsilon\chi_j f_{ij}}\right)^{\frac{1}{\varepsilon-1}}\right)^{\theta}\chi_j f_{ij} \\ &= \sum_{i=1}^{J}\mu_j^{\theta}\left[\frac{X_i}{\varepsilon}\left(\frac{\varepsilon\tau_{ij}\chi_j}{\varepsilon-1}\right)^{1-\varepsilon}\right]^{\frac{\theta}{\varepsilon-1}}\left(P_i\right)^{\theta}\left(\chi_j f_{ij}\right)^{1-\frac{\theta}{\varepsilon-1}}\frac{\varepsilon-1}{\theta-(\varepsilon-1)} \\ f_e &= \sum_{i=1}^{J}\mu_j^{\theta}\left[\frac{X_i}{\varepsilon}\left(\frac{\varepsilon\tau_{ij}}{\varepsilon-1}\right)^{1-\varepsilon}\right]^{\frac{\theta}{\varepsilon-1}}\left(P_i\right)^{\theta}\left(f_{ij}\right)^{1-\frac{\theta}{\varepsilon-1}}\left(\chi_j\right)^{-\frac{\theta\varepsilon}{\varepsilon-1}}\frac{\varepsilon-1}{\theta-(\varepsilon-1)}. \end{aligned}$$

Re-arrange:

$$\left(\frac{1}{\varepsilon}\right)^{-\frac{\theta}{\varepsilon-1}}\left(\frac{\varepsilon}{\varepsilon-1}\right)^{\theta}\left(\frac{\theta-(\varepsilon-1)}{\varepsilon-1}\right)\mu_j^{-\theta}\left(\chi_j\right)^{\frac{\theta\varepsilon}{\varepsilon-1}} f_e=\sum_{i=1}^{J}\left[X_i\left(\tau_{ij}\right)^{1-\varepsilon}\right]^{\frac{\theta}{\varepsilon-1}}\left(f_{ij}\right)^{1-\frac{\theta}{\varepsilon-1}}\left(P_i\right)^{\theta}$$
$$\left(\frac{1}{\varepsilon}\right)^{-\frac{\theta}{\varepsilon-1}}\left(\frac{\varepsilon}{\varepsilon-1}\right)^{\theta}\left(\frac{\theta-(\varepsilon-1)}{\varepsilon-1}\right)\mu_j^{-\theta}\left(\chi_j\right)^{\frac{\theta\varepsilon}{\varepsilon-1}} f_e=\sum_{i=1}^{J}\left(X_i\right)^{\frac{\theta}{\varepsilon-1}}\left(P_i\right)^{\theta}\Upsilon_{ij}.$$

The above equation, for all the regions $j=1,\cdots,J$, can be written in matrix form:

$$\begin{bmatrix}\Upsilon_{11}\left(X_1\right)^{\frac{\theta}{\varepsilon-1}} & \Upsilon_{21}\left(X_2\right)^{\frac{\theta}{\varepsilon-1}} & \cdots\Upsilon_{J1}\left(X_J\right)^{\frac{\theta}{\varepsilon-1}}\\ \vdots & \vdots & \vdots\\ \Upsilon_{1J}\left(X_1\right)^{\frac{\theta}{\varepsilon-1}} & \Upsilon_{2J}\left(X_2\right)^{\frac{\theta}{\varepsilon-1}} & \cdots\Upsilon_{JJ}\left(X_J\right)^{\frac{\theta}{\varepsilon-1}}\end{bmatrix}\begin{bmatrix}\left(P_1\right)^{\theta}\\ \vdots\\ \left(P_J\right)^{\theta}\end{bmatrix}=\begin{bmatrix}\left(\frac{1}{\varepsilon}\right)^{-\frac{\theta}{\varepsilon-1}}\left(\frac{\varepsilon}{\varepsilon-1}\right)^{\theta}\left(\frac{\theta-(\varepsilon-1)}{\varepsilon-1}\right)\mu_1^{-\theta}\left(\chi_1\right)^{\frac{\theta\varepsilon}{\varepsilon-1}} f_e\\ \vdots\\ \left(\frac{1}{\varepsilon}\right)^{-\frac{\theta}{\varepsilon-1}}\left(\frac{\varepsilon}{\varepsilon-1}\right)^{\theta}\left(\frac{\theta-(\varepsilon-1)}{\varepsilon-1}\right)\mu_J^{-\theta}\left(\chi_J\right)^{\frac{\theta\varepsilon}{\varepsilon-1}} f_e\end{bmatrix}.$$

Denote the LHS matrix as AA and the RHS vector as BB; the above equation provides a solution to the vector $\left(P_j\right)^{\theta}$:

$$\left(P_j\right)^{\theta}=AA^{-1}*BB$$

Note that from equation (B.4), we have another solution of price, which we denote as $\left(P_j\right)^{\theta}=DD$. Combining the two solutions, it is straightforward to see $BB=AA*DD$:

$$\begin{bmatrix}\left(\frac{1}{\varepsilon}\right)^{-\frac{\theta}{\varepsilon-1}}\left(\frac{\varepsilon}{\varepsilon-1}\right)^{\theta}\left(\frac{\theta-(\varepsilon-1)}{\varepsilon-1}\right)\mu_1^{-\theta}\left(\chi_1\right)^{\frac{\theta\varepsilon}{\varepsilon-1}} f_e\\ \vdots\\ \left(\frac{1}{\varepsilon}\right)^{-\frac{\theta}{\varepsilon-1}}\left(\frac{\varepsilon}{\varepsilon-1}\right)^{\theta}\left(\frac{\theta-(\varepsilon-1)}{\varepsilon-1}\right)\mu_J^{-\theta}\left(\chi_J\right)^{\frac{\theta\varepsilon}{\varepsilon-1}} f_e\end{bmatrix}=\begin{bmatrix}\Upsilon_{11}\left(X_1\right)^{\frac{\theta}{\varepsilon-1}} & \Upsilon_{21}\left(X_2\right)^{\frac{\theta}{\varepsilon-1}} & \cdots\Upsilon_{J1}\left(X_J\right)^{\frac{\theta}{\varepsilon-1}}\\ \vdots & \vdots & \vdots\\ \Upsilon_{1J}\left(X_1\right)^{\frac{\theta}{\varepsilon-1}} & \Upsilon_{2J}\left(X_2\right)^{\frac{\theta}{\varepsilon-1}} & \cdots\Upsilon_{JJ}\left(X_J\right)^{\frac{\theta}{\varepsilon-1}}\end{bmatrix}$$
$$*\begin{bmatrix}\left(\frac{\varepsilon}{\varepsilon-1}\right)^{\theta}\left(\frac{\theta}{\theta-(\varepsilon-1)}\right)^{-1}\left(\frac{X_1}{\varepsilon}\right)^{-\frac{\theta-(\varepsilon-1)}{(\varepsilon-1)}}\left[\sum_{i=1}^{J} I_i\left(\Upsilon_{1i}\right)\left(\mu_i\right)^{\theta}\left(\chi_i\right)^{-\theta-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}\right]^{-1}\\ \vdots\\ \left(\frac{\varepsilon}{\varepsilon-1}\right)^{\theta}\left(\frac{\theta}{\theta-(\varepsilon-1)}\right)^{-1}\left(\frac{X_J}{\varepsilon}\right)^{-\frac{\theta-(\varepsilon-1)}{(\varepsilon-1)}}\left[\sum_{i=1}^{J} I_i\left(\Upsilon_{Ji}\right)\left(\mu_i\right)^{\theta}\left(\chi_i\right)^{-\theta-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}\right]^{-1}\end{bmatrix}$$

After some manipulation and simplification:

$$\begin{bmatrix} (\mu_1)^{-\theta} & 0 & \cdots & 0 \\ 0 & (\mu_2)^{-\theta} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (\mu_J)^{-\theta} \end{bmatrix} \begin{bmatrix} \frac{\theta\varepsilon}{\varepsilon-1} (\chi_1)^{\frac{\theta\varepsilon}{\varepsilon-1}} f_e \\ \vdots \\ \frac{\theta\varepsilon}{\varepsilon-1} (\chi_J)^{\frac{\theta\varepsilon}{\varepsilon-1}} f_e \end{bmatrix} = \begin{bmatrix} \Upsilon_{11} & \Upsilon_{21} & \cdots \Upsilon_{J1} \\ \vdots & \vdots & \vdots \\ \Upsilon_{1J} & \Upsilon_{2J} & \cdots \Upsilon_{JJ} \end{bmatrix}$$

$$* \begin{bmatrix} (X_1)^{\frac{\theta}{\varepsilon-1}} & 0 & \cdots & 0 \\ 0 & (X_2)^{\frac{\theta}{\varepsilon-1}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (X_J)^{\frac{\theta}{\varepsilon-1}} \end{bmatrix} * \begin{bmatrix} (X_1)^{-\frac{\theta-(\varepsilon-1)}{(\varepsilon-1)}} \left[\sum_{i=1}^{J} I_i (\Upsilon_{1i}) (\mu_i)^{\theta} (\chi_i)^{-\theta-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}\right]^{-1} \\ \vdots \\ (X_J)^{-\frac{\theta-(\varepsilon-1)}{(\varepsilon-1)}} \left[\sum_{i=1}^{J} I_i (\Upsilon_{Ji}) (\mu_i)^{\theta} (\chi_i)^{-\theta-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}\right]^{-1} \end{bmatrix}$$

$$\begin{bmatrix} \frac{\theta\varepsilon}{\varepsilon-1} (\chi_1)^{\frac{\theta\varepsilon}{\varepsilon-1}} f_e \\ \vdots \\ \frac{\theta\varepsilon}{\varepsilon-1} (\chi_J)^{\frac{\theta\varepsilon}{\varepsilon-1}} f_e \end{bmatrix} = \begin{bmatrix} \Upsilon_{11} (\mu_1)^{\theta} & \Upsilon_{21} (\mu_1)^{\theta} & \cdots \Upsilon_{J1} (\mu_1)^{\theta} \\ \vdots & \vdots & \vdots \\ \Upsilon_{1J} (\mu_J)^{\theta} & \Upsilon_{2J} (\mu_J)^{\theta} & \cdots \Upsilon_{JJ} (\mu_J)^{\theta} \end{bmatrix} * \begin{bmatrix} \frac{X_1}{\sum_{i=1}^{J} I_i (\Upsilon_{1i}) (\mu_i)^{\theta} (\chi_i)^{-\theta-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}} \\ \vdots \\ \frac{X_J}{\sum_{i=1}^{J} I_i (\Upsilon_{Ji}) (\mu_i)^{\theta} (\chi_i)^{-\theta-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}} \end{bmatrix}$$

Pre-multiply both sides of the equation with the diagonal matrix $(\chi_j)^{-\frac{\theta\varepsilon-\varepsilon+1}{\varepsilon-1}}$:

$$\begin{bmatrix} \frac{\theta\varepsilon}{\varepsilon-1} \chi_1 f_e \\ \vdots \\ \frac{\theta\varepsilon}{\varepsilon-1} \chi_J f_e \end{bmatrix} = \begin{bmatrix} \Upsilon_{11} (\mu_1)^{\theta} (\chi_1)^{-\frac{\theta\varepsilon-\varepsilon+1}{\varepsilon-1}} & \Upsilon_{21} (\mu_1)^{\theta} (\chi_1)^{-\frac{\theta\varepsilon-\varepsilon+1}{\varepsilon-1}} & \cdots \Upsilon_{J1} (\mu_1)^{\theta} (\chi_1)^{-\frac{\theta\varepsilon-\varepsilon+1}{\varepsilon-1}} \\ \vdots & \vdots & \vdots \\ \Upsilon_{1J} (\mu_J)^{\theta} (\chi_J)^{-\frac{\theta\varepsilon-\varepsilon+1}{\varepsilon-1}} & \Upsilon_{2J} (\mu_J)^{\theta} (\chi_J)^{-\frac{\theta\varepsilon-\varepsilon+1}{\varepsilon-1}} & \cdots \Upsilon_{JJ} (\mu_J)^{\theta} (\chi_J)^{-\frac{\theta\varepsilon-\varepsilon+1}{\varepsilon-1}} \end{bmatrix}$$
$$* \begin{bmatrix} (X_1) \left[\sum_{i=1}^{J} I_i (\Upsilon_{1i}) (\mu_i)^{\theta} (\chi_i)^{-\theta-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}\right]^{-1} \\ \vdots \\ (X_J) \left[\sum_{i=1}^{J} I_i (\Upsilon_{Ji}) (\mu_i)^{\theta} (\chi_i)^{-\theta-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}\right]^{-1} \end{bmatrix}.$$

Denoting the RHS matrix on the first line with elements $\Upsilon_{ij} (\mu_j)^{\theta} (\chi_j)^{-\frac{\theta\varepsilon-\varepsilon+1}{\varepsilon-1}}$ as Ψ, we can re-write the above equation as

$$\Psi^{-1} * \begin{bmatrix} \frac{\theta\varepsilon}{\varepsilon-1} \chi_1 f_e \\ \vdots \\ \frac{\theta\varepsilon}{\varepsilon-1} \chi_J f_e \end{bmatrix} = \begin{bmatrix} (X_1) \left[\sum_{i=1}^{J} I_i (\Upsilon_{1i}) (\mu_i)^{\theta} (\chi_i)^{-\theta-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}\right]^{-1} \\ \vdots \\ (X_J) \left[\sum_{i=1}^{J} I_i (\Upsilon_{Ji}) (\mu_i)^{\theta} (\chi_i)^{-\theta-\frac{\theta-(\varepsilon-1)}{\varepsilon-1}}\right]^{-1} \end{bmatrix}.$$

Denote the LHS vector as

$$\zeta = \Psi^{-1} * \begin{bmatrix} \frac{\theta\varepsilon}{\varepsilon-1}\chi_1 f_e \\ \vdots \\ \frac{\theta\varepsilon}{\varepsilon-1}\chi_J f_e \end{bmatrix}.$$

It is straightforward to see, with the understanding that ζ_j is the j-th element of vector ζ:

$$\begin{aligned}
\begin{bmatrix} \frac{X_1}{\zeta_1} \\ \vdots \\ \frac{X_J}{\zeta_J} \end{bmatrix} &= \begin{bmatrix} \sum_{i=1}^{J} I_i \left(\Upsilon_{1i}\right) \left(\mu_i\right)^{\theta} \left(\chi_i\right)^{-\theta - \frac{\theta-(\varepsilon-1)}{\varepsilon-1}} \\ \vdots \\ \sum_{i=1}^{J} I_i \left(\Upsilon_{Ji}\right) \left(\mu_i\right)^{\theta} \left(\chi_i\right)^{-\theta - \frac{\theta-(\varepsilon-1)}{\varepsilon-1}} \end{bmatrix} \\
&= \begin{bmatrix} \Upsilon_{11} \left(\mu_1\right)^{\theta} \left(\chi_1\right)^{-\frac{\theta\varepsilon-\varepsilon+1}{\varepsilon-1}} & \Upsilon_{12} \left(\mu_2\right)^{\theta} \left(\chi_2\right)^{-\frac{\theta\varepsilon-\varepsilon+1}{\varepsilon-1}} & \cdots \Upsilon_{1J} \left(\mu_J\right)^{\theta} \left(\chi_J\right)^{-\frac{\theta\varepsilon-\varepsilon+1}{\varepsilon-1}} \\ \vdots & \vdots & \vdots \\ \Upsilon_{J1} \left(\mu_1\right)^{\theta} \left(\chi_1\right)^{-\frac{\theta\varepsilon-\varepsilon+1}{\varepsilon-1}} & \Upsilon_{J2} \left(\mu_2\right)^{\theta} \left(\chi_2\right)^{-\frac{\theta\varepsilon-\varepsilon+1}{\varepsilon-1}} & \cdots \Upsilon_{JJ} \left(\mu_J\right)^{\theta} \left(\chi_J\right)^{-\frac{\theta\varepsilon-\varepsilon+1}{\varepsilon-1}} \end{bmatrix} \begin{bmatrix} I_1 \\ \vdots \\ I_J \end{bmatrix} \\
&= \Psi' \begin{bmatrix} I_1 \\ \vdots \\ I_J \end{bmatrix}.
\end{aligned}$$

From the last line the solution of the vector I_j follows:

$$\begin{bmatrix} I_1 \\ \vdots \\ I_J \end{bmatrix} = \left(\Psi'\right)^{-1} \begin{bmatrix} \frac{X_1}{\zeta_1} \\ \vdots \\ \frac{X_J}{\zeta_J} \end{bmatrix}. \tag{B.10}$$

B.6 Updating N_j

N_j is directly updated using equation (17), conditional on the solution of w_j, P_j, and I_j.

B.7 Fixed Entry

Profits In the special case of "fixed entry", we set the mass of entrants in each region to an exogenous level, denoted as $\bar{I}_j$. In this case, the firms earn profits, which will be distributed back to all the residents living in region j, including immigrants. The profit of a firm originating in j

and selling to i with a productivity a is:

$$\pi_{ij}(a) - \chi_j f_{ij}.$$

At the aggregate level, denote the total profit of firms selling from j to i as Ξ_{ij}:

$$\begin{aligned}
\Xi_{ij} &= \bar{I}_j \left[\int_0^{a_{ij}} \pi_{ij}(a) dG(a) - \chi_j f_{ij} G_j(a_{ij}) \right] \\
&= \bar{I}_j \frac{1}{\varepsilon} \frac{X_i}{(P_i)^{1-\varepsilon}} \left(\frac{\varepsilon}{\varepsilon - 1} \tau_{ij} \chi_j \right)^{1-\varepsilon} \int_0^{a_{ij}} (a(k))^{1-\varepsilon} dG(a) - \bar{I}_j \chi_j f_{ij} G_j(a_{ij}) \\
&= \frac{1}{\varepsilon} X_{ij} - \bar{I}_j \chi_j f_{ij} \mu_j^{\theta} \left[\frac{\varepsilon - 1}{\varepsilon} \frac{P_i}{\tau_{ij}\chi_j} \left(\frac{X_i}{\varepsilon \chi_j f_{ij}} \right)^{\frac{1}{\varepsilon-1}} \right]^{\theta} \\
&= \frac{1}{\varepsilon} X_{ij} - \bar{I}_j \mu_j^{\theta} (X_i)^{\frac{\theta}{\varepsilon-1}} (P_i)^{\theta} \left(\frac{\varepsilon - 1}{\varepsilon} \right)^{\theta} (\tau_{ij})^{-\theta} (\chi_j)^{\frac{\theta\varepsilon}{1-\varepsilon}+1} (f_{ij})^{\frac{\theta}{1-\varepsilon}+1} \varepsilon^{-\frac{\theta}{\varepsilon-1}} \\
&= \frac{1}{\varepsilon} X_{ij} - \frac{\theta - (\varepsilon - 1)}{\theta\varepsilon} X_{ij} \\
&= \frac{\varepsilon - 1}{\theta\varepsilon} X_{ij}.
\end{aligned}$$

Denote the aggregate profit in region j as Ξ_j, it is then straightforward to see that the aggregate profit must be a constant share of the total sales:

$$\Xi_j = \sum_{i=1}^{J} \Xi_{ij} = \frac{\varepsilon - 1}{\theta\varepsilon} \sum_{i=1}^{J} X_{ij} = \frac{\varepsilon - 1}{\theta\varepsilon} X_j.$$

Expenditure The expenditure on the differentiated goods in the urban region adopts a new expression as well:

$$X_j = \gamma \left[(w_j + T_c) N_j + \Xi_j \right] + (1 - \beta) \left(1 - \frac{\varepsilon - 1}{\theta\varepsilon} \right) X_j.$$

Different from the expression in the baseline model, the total income in the urban region becomes $(w_j + T_c)N_j + \Xi_j$. Similarly, the expenditure on intermediate goods is now $(1-\beta)\left(1 - \frac{\varepsilon-1}{\theta\varepsilon}\right) X_j$, taking into account that $\left(1 - \frac{\varepsilon-1}{\theta\varepsilon}\right) X_j$ is the aggregate costs of all the firms in region j. Simplify the goods market clearing condition:

$$
\begin{aligned}
X_j &= \gamma \left[(w_j + T_c)N_j + \frac{\varepsilon - 1}{\theta\varepsilon} X_j\right] + (1-\beta)\left(1 - \frac{\varepsilon - 1}{\theta\varepsilon}\right) X_j \\
&= \gamma(w_j + T_c)N_j + \left[1 - \beta + \frac{\varepsilon - 1}{\theta\varepsilon}\left(\gamma - (1-\beta)\right)\right] X_j \\
\left[\beta - \frac{\varepsilon - 1}{\theta\varepsilon}\left(\gamma - (1-\beta)\right)\right] X_j &= \gamma(w_j + T_c)N_j,
\end{aligned}
$$

which leads to

$$
X_j = \frac{\gamma}{\beta - \frac{\varepsilon-1}{\theta\varepsilon}\left(\gamma - (1-\beta)\right)}(w_j + T_c)N_j. \tag{B.11}
$$

Note that the above equation implies a parameter restriction that $\beta - \frac{\varepsilon-1}{\theta\varepsilon}\left(\gamma - (1-\beta)\right) > 0$, which is met in all the specifications in the paper.

Income Taking the expression of total expenditure in equation (B.11), the total income in the j becomes:

$$
\begin{aligned}
(w_j + T_c)N_j + \Xi_j &= (w_j + T_c)N_j + \frac{\varepsilon - 1}{\theta\varepsilon} X_j \\
&= (w_j + T_c)N_j + \frac{\varepsilon - 1}{\theta\varepsilon} \frac{\gamma}{\beta - \frac{\varepsilon-1}{\theta\varepsilon}\left(\gamma - (1-\beta)\right)}(w_j + T_c)N_j \\
&= \left[1 + \frac{\varepsilon - 1}{\theta\varepsilon} \frac{\gamma}{\beta - \frac{\varepsilon-1}{\theta\varepsilon}\left(\gamma - (1-\beta)\right)}\right](w_j + T_c)N_j \\
&= \left[\frac{\beta + \frac{\varepsilon-1}{\theta\varepsilon}(1-\beta)}{\beta + \frac{\varepsilon-1}{\theta\varepsilon}(1-\beta) - \gamma\frac{\varepsilon-1}{\theta\varepsilon}}\right](w_j + T_c)N_j \\
&= \rho(w_j + T_c)N_j,
\end{aligned}
$$

where

$$
\rho = \frac{\beta + \frac{\varepsilon-1}{\theta\varepsilon}(1-\beta)}{\beta + \frac{\varepsilon-1}{\theta\varepsilon}(1-\beta) - \gamma\frac{\varepsilon-1}{\theta\varepsilon}} > 1.
$$

Land Rents Lastly, the aggregate land rent is now computed as:

$$T_c = \frac{R_{A,c}L_{A,c} + (1-\alpha-\gamma)\left[\sum_{i\in J_c\cap\mathcal{U}}\left[(w_i+T_c)\,N_i+\Xi_i\right]+\sum_{i\in J_c\cap\mathcal{R}}(w_i+T_c)\,N_i\right]}{\bar{N}_c}.$$

Simplify the expression, and use j to index the rural region in country c:

$$\begin{aligned}
T_c &= \frac{\eta\nu^{-1}w_jN_j + (1-\alpha-\gamma)\left[\sum_{i\in J_c\cap\mathcal{U}}\rho\,(w_i+T_c)\,N_i+(w_j+T_c)\,N_j\right]}{\bar{N}_c} \\
&= \frac{\eta\nu^{-1}w_jN_j + (1-\alpha-\gamma)\left[\sum_{i\in J_c\cap\mathcal{U}}\rho w_iN_i+w_jN_j\right]}{\bar{N}_c} + (1-\alpha-\gamma)\,T_c\frac{\sum_{i\in J_c\cap\mathcal{U}}\rho N_i+N_j}{\bar{N}_c} \\
&= \frac{\eta\nu^{-1}w_jN_j + (1-\alpha-\gamma)\left[\sum_{i\in J_c\cap\mathcal{U}}\rho w_iN_i+w_jN_j\right]}{\left[1-(1-\alpha-\gamma)\,\frac{\sum_{i\in J_c\cap\mathcal{U}}\rho N_i+N_j}{\bar{N}_c}\right]\bar{N}_c}.
\end{aligned}$$

Urban Wage The algorithm to solve the urban wage rates is not affected. To see this, first note that the urban income accounting becomes:

$$(w_j+T_c)N_j+\Xi_j = \sum_{i\in\mathcal{R}}X_{ij}+\sum_{i\in\mathcal{U}}X_{ij}-(1-\beta)\left(1-\frac{\varepsilon-1}{\theta\varepsilon}\right)X_j+(1-\alpha-\gamma)\left[(w_j+T_c)N_j+\Xi_i\right].$$

In the expression above, the LHS is the total income in j, and the RHS is the income source. The first part is the value-added from the differentiated sector, which is the sales to all the urban and the rural regions, minus the costs of the intermediate products. The second part of the land rent. Similar to the baseline model, trade balance with the rural regions implies $\sum_{i\in\mathcal{R}}X_{ij} = \alpha[(w_j+T_c)N_j+\Xi_j]$, which leads to:

$$\gamma\left[(w_j+T_c)N_j+\Xi_j\right]+(1-\beta)\left(1-\frac{\varepsilon-1}{\theta\varepsilon}\right)X_j=\sum_{i\in\mathcal{U}}X_{ij}.$$

Substitute in the expression of Ξ_j and X_j from equation (B.11):

$$\begin{aligned}
\gamma\rho(w_j+T_c)N_j+(1-\beta)\left(1-\frac{\varepsilon-1}{\theta\varepsilon}\right)\frac{\gamma}{\beta-\frac{\varepsilon-1}{\theta\varepsilon}\left(\gamma-(1-\beta)\right)}(w_j+T_c)N_j &= \sum_{i\in\mathcal{U}}X_{ij} \\
\frac{\gamma}{\beta-\frac{\varepsilon-1}{\theta\varepsilon}\left(\gamma-(1-\beta)\right)}(w_j+T_c)N_j &= \sum_{i\in\mathcal{U}}X_{ij}.
\end{aligned}$$

Substitute in the expression of X_{ij} from equation (B.6), we arrive at the same solution as in the baseline model as in equation (B.7).

Rural Wage If country c engages in international trade in the agricultural products, then its rural wage rate is still implicitly pinned down by equation (B.9), the same as in the baseline model. In the case of agricultural autarky, the rural wage rate, w_j, is pined down by the modified market clearing condition:

$$w_j N_j = \nu \left[\alpha \left(\sum_{i \in J_c \cap \mathcal{U}} \rho \left(w_i + T_c \right) N_i \right) + \alpha (w_j + T_c) N_j \right]$$
$$\left(1 - \alpha\nu\right) w_j N_j = \alpha\nu T_c N_j + \nu \left[\alpha \sum_{i \in J_c \cap \mathcal{U}} \rho \left(w_i + T_c \right) N_i \right].$$

Substitute in the modified expression of T_c:

$$\begin{aligned}
\frac{1-\alpha\nu}{\alpha\nu} w_j N_j &= T_c N_j + \sum_{i \in J_c \cap \mathcal{U}} \rho \left(w_i + T_c \right) N_i \\
&= T_c \left(N_j + \rho \sum_{i \in J_c \cap \mathcal{U}} N_i \right) + \sum_{i \in J_c \cap \mathcal{U}} \rho w_i N_i \\
&= \frac{\eta \nu^{-1} w_j N_j + \left(1 - \alpha - \gamma\right) \left[\sum_{i \in J_c \cap \mathcal{U}} \rho w_i N_i + w_j N_j \right]}{\bar{N}_c - \left(1 - \alpha - \gamma\right) \left(\sum_{i \in J_c \cap \mathcal{U}} \rho N_i + N_j \right)} \left(N_j + \rho \sum_{i \in J_c \cap \mathcal{U}} N_i \right) + \sum_{i \in J_c \cap \mathcal{U}} \rho w_i N_i \\
&= Z_c \left\{ \eta \nu^{-1} w_j N_j + \left(1 - \alpha - \gamma\right) \left[\sum_{i \in J_c \cap \mathcal{U}} \rho w_i N_i + w_j N_j \right] \right\} + \sum_{i \in J_c \cap \mathcal{U}} \rho w_i N_i
\end{aligned}$$

where:

$$Z_c = \frac{N_j + \rho \sum_{i \in J_c \cap \mathcal{U}} N_i}{\bar{N}_c - \left(1 - \alpha - \gamma\right) \left(N_j + \rho \sum_{i \in J_c \cap \mathcal{U}} N_i \right)}.$$

Simplify the solution:

$$\frac{1-\alpha\nu}{\alpha\nu} w_j N_j = Z_c \left(\frac{\eta}{\nu} + 1 - \alpha - \gamma \right) w_j N_j + \left(Z_c \left(1 - \alpha - \gamma\right) + 1 \right) \sum_{i \in J_c \cap \mathcal{U}} \rho w_i N_i$$
$$w_j N_j = \frac{Z_c \left(1 - \alpha - \gamma\right) + 1}{\frac{1-\alpha\nu}{\alpha\nu} - Z_c \left(\frac{\eta}{\nu} + 1 - \alpha - \gamma \right)} \sum_{i \in J_c \cap \mathcal{U}} \rho w_i N_i.$$

Price Index and Trade Flow The expressions in these parts are not affected by shutting down firm entry.

C Data and Quantification

This appendix provides the details regarding the data sources and the quantification of the model. We organize the discussion by data source.

C.1 Data Sources, Global

The World Development Indicators We use several components of the WDI. For the following variables, we take the average value between 2000 and 2005 for the equilibrium in the year 2005, and the average between 2010 and 2015 for the equilibrium in the year 2015:

- The employment in agriculture variable (SL.AGR.EMPL.ZS) is used to infer the rural population.
- The cereal production data (AG.PRD.CREL.MT) is used to infer the agriculture productivity.
- The time required to start a business variable (IC.REG.DURS) is used to infer the fixed costs of operation, f_i.

The Penn World Table We use the 9.1 version of the PWT in this paper. Our measure of population (pop) comes from the PWT. We use the average population between 2000 and 2005 for the 2005 calibration, and the average between 2010 and 2015 for the 2015 calibration.

The differentiation between the ROW and the ODC is based on the per capita GDP, which we define as "$rgdpo/pop$", averaged between 2000 and 2015. A country with average per capita GDP less than 2/3 of the USA is defined as ODC.

The cross-sectional TFP used to calibrate urban productivity is the variable "$ctfp$", and the inter-temporal TFP used to calibrate the growth of urban productivity between 2005 and 2015 is "$rtfpna$".

The OECD Inter-Country Input-Output Tables We use the 2018 version of the ICIO tables to infer the bilateral trade flow matrix between the three countries, which is in turn used to compute the variable trade costs. The 2018 version provides annual data from the year 2005 to 2015; we use the data from respective years for our year-specific calibration of τ_{ij}.

The ESCAP-World Bank Trade Costs Database We use this database for two purposes. In the first, we use this to infer $\bar{\tau}$, the ratio of agriculture trade costs to manufacturing trade costs. We restrict the sample to the year 2005, and restrict the reporting countries and the partner ones to be within our sample as listed in Table A.2. Using the variable names from the dataset, we compute $\bar{\tau}$ as the simple average of $tij(AB)/tij(D)$ across all observations, where tij corresponds to our variable trade cost minus 1 and AB refers to the agricultural sector, D to the manufacturing sector.

We also use this dataset to compute the change in the trade barrier of China over time. The trade costs measures are symmetric and therefore the trade barrier refers to both the inbound and the outbound barrier. We compute the simple average across all trading partners across all sectors. The average iceberg cost of selling into China was 3.605 in 1996, and it declined by 5.1% to $(3.605 - 1) * (1 - 0.051) + 1 = 3.471$ in 2006.

The USDA-ERS Database We use the data for three purposes: to calibrate the production function of each country, to compute the land endowment, and to compute the rural productivity. We use the 2019 Oct 1st version of the data that covers 187 countries between 1961 and 2016. The land endowment data are at the yearly frequency so we use the respective years for the 2005 and 2015 calibration. The factor-share data come at the decade frequency, so we use the factor share in 2000–2010 for the calibration of 2005, and 2010–2020 for the 2015 calibration.

C.2 Data Sources, China

Input-Output Table of China We use the 2002 Input-Output Table of China to estimate the agriculture share in consumption (α) and the labor share in differentiated products (β). The agriculture consumption share is computed as THC(1), and the total consumption is computed as $\sum_{i=1}^{42}$ THC(i). The labor share is the summation of all the value-added terms (TVA); we define industries 02 to 21 as the differentiated industries.

One-Percent Population Survey The One-Percent Population Survey was conducted in 2005 and 2015 by the National Statistics Bureau of China. Our sample in the year 2005 contains 2.6 million individuals, and in 2015, 1.4 million. We estimate the migration probability matrix using this data.

We identify the original location of the individual as the follows. If the individual reported a rural hukou in the 2005 survey (Question 11), or was entitled to contract rural land (*Tu Di Cheng Bao*) in the 2015 survey (Question 11), then the individual is classified as originating from the rural region by both definitions of a migrant (hukou-migrant or five-year-migrant). The original prefecture for a hukou-migrant is the place of hukou registration.

The current prefecture of the individual is readily available in the survey. To distinguish between rural and urban areas, we rely on the "Urban-Rural Codes" (*Cheng Xiang Hua Fen Ma*) reported in the survey. We classify the following codes as urban: 111 (city center, *Shi Zhong Xin*), 112 (city suburb, *Cheng Xiang Jie He Bu*), 121 (town center, *Zhen Zhong Xin*), 122 (township suburbs, *Zhen Xiang Jie He Bu*), and the following codes as rural: 210 (large village, *Xiang*) and 220 (village, *Cun*).

We use the weighted population count in the surveys to account for the sampling weights, and compute the out-migration probability from region j to region i as the sum of population weights that move from j to i divided by the sum of the original population weights of region j.

Economic Census The Economic Census is used to compute the firm-to-population ratio in China, which is in turn used to calibrate f_e. We use the *First Economic Census (2004)* for the calibration in 2005, and the *Third Economic Census (2013)* for 2015. We define firms as "legal entity (*Fa Ren*)".

Population Census The Population Censuses in 2000 and 2010 are used to construct the initial population distributions in the 2005 and 2015 calibrations, respectively. As mentioned in Section 4.2.1, the relative population between the MUR and OUR is needed. According to our definitions of cities and the two urban regions, the urban population (*Shi Xia Qu Ren Kou*) from the Population Censuses is used to calculate the population ratio between the MUR and OUR.

City Statistical Yearbooks We use the *City Statistical Yearbooks* to construct the GDP at the city level, which was then used in many parts of the calibration exercise, such as the estimation of productivity, amenity, and $\bar{R}_i$. To be consistent with our definition of cities, the urban GDP of a prefecture is defined as the sum of the secondary and tertiary GDP in the urban districts of that prefecture (*Shi Xia Qu*).

In addition, we also use the *City Statistical Yearbooks* to estimate the city-level amenity. The following variables in the vector $\mathbf{X}_i$ come from the *City Statistical Yearbooks*: the number of universities, middle schools, and primary schools; the number of university, middle school, and primary school teachers; the number of public library books; the number of hospitals, hospital beds, and doctors; and the percentage of green fields in constructed areas.

City-Level Climate and Geographical Variables The city-level temperature and precipitation data come from the National Oceanic and Atmospheric Administration (NOAA). We measure the city-level climate using the 0.5-degree cell in which the city center resides. The elevation of a city comes from the GTOPO30 database, and the slope is inferred from the elevation data. Lastly, the ease of access to the national transportation network comes from Ma and Tang (2020).

CBD Land Rents A 2004 ordinance requires that land sales by Chinese governments at all levels must be publicized on the internet. However, only after 2007 did such data become complete and relatively organized on government websites. We have land sales data from 2007 to 2017. As the data in 2007 is still relatively sparse, we pool the data in both 2007 and 2008 to proxy for 2005. Correspondingly, we use the data in 2017 to proxy for 2015. To proxy the CBD land rent, we first use the average price of the top 10% land sales prices, and then annualize this according to the number of years of the leasehold and a 10% interest rate. The 2017 data is quite clean, but there are unreasonable outliers in the 2007–08 data. In some small cities, some annualized land sale prices in 2007–08 are even much higher than the so-calculated CBD prices in Shenzhen and Beijing in 2017. When all of the annualized land sale prices in the 2007–08 data are ranked, we find that the first possibly sensible highest price is Guangzhou's highest price at 13108, which is still higher than Bejing's average of the top 3% prices in 2017. Hence, we use 13108 as a cutoff to trim all of the higher prices to alleviate concerns over measurement errors.

Baidu Commuting Data Baidu Maps publishes annual reports on urban transportation. There is a specific table on commuting distance and time for a hundred selected cities. For further details, see `https://jiaotong.baidu.com/reports/`.

C.3 Estimation of μ_j at the city level

Ma and Tang (2020) estimate city-level productivity in a heterogeneous-firm model setup similar to the model in this paper. They back out the city-level productivity from the residual of the following regression:

$$\log(w_j) = b_0 + b_1 \log(N_j) + b_2 \log(\mathbf{MA}_j) + \nu_j,$$

where N_j is the population of city j and w_j is approximated by the per capita GDP of the city. According to our definition of cities, the city population is the population in the collection of districts in a prefecture (*Shi Xia Qu Ren Kou*), and the city GDP is the sum of secondary and tertiary GDP in these districts; both variables are obtained from China City Statistical Yearbooks. Here, term MA_j summarizes the market access from location j that encompasses the internal transportation network and market size distribution in China. Following Donaldson and Hornbeck (2016), the first-order approximation of MA_j can be written as

$$\text{MA}_j = \sum_{i=1}^{J} w_i N_i (\tau_{ij})^{-\theta},$$

where θ is the trade elasticity. This term captures the ease of access to markets given a trade cost matrix $\{\tau_{ij}\}$. The trade cost matrix is obtained from Ma and Tang (2020). The city-level productivity is then computed as $\mu_j = \exp(\tilde{\nu}_j/\theta)$, where $\tilde{\nu}_j$ is the residual of the above regression.[1] The city-level productivities are then used to infer the region-level productivities.

Lastly, note that the above regression excludes foreign economies. The exclusion is due to two reasons. The first reason is data limitations: data on internal trade costs (τ_{ij}) in China is scarce, and the most detailed matrix from earlier work lacks information on trade costs with foreign economies.

[1] We cannot directly use the estimated city-level productivity from Ma and Tang (2020) as their paper uses a different trade elasticity.

The second reason is inconsistency in the unit of observation. Whereas the data points within China are at the city level, foreign economies in this regression would have been countries or even groups of countries. For this reason, foreign economies in this regression would be much larger in size than the cities in China, and they distort the point estimates and the residuals as commonly seen in an OLS setting. For these two reasons, we include only the cities in China in the reduced-form regression.

D Robustness Checks

In this section, we present four robustness checks. In the first, we use a higher migration elasticity, κ, and in the second one, a higher ε to capture a world with weaker market power. In the third exercise, we experiment with a higher expenditure share of land consumption, and in the last, we shut down the channel of entry and exit of firms. In all the exercises, we re-calibrate the migration frictions and the fixed costs of entry, and report these parameters in Table D.1. The main welfare results are reported in Tables D.2 and D.3, and the migration probabilities in Table D.4.

D.1 Higher Migration Elasticity

In the baseline quantification of the model we use a migration elasticity of $\kappa = 1.63$. Although our choice of κ lies within the range of common estimates between 1.4 and 3.3 in the literature, it nevertheless is closer to the lower end. As a robustness check, we re-calibrate $\{\lambda_{ij}\}$ and f_e in the year 2015 using $\kappa = 3.3$ from Monte et al. (2018), the estimate on the higher end.

A higher migration elasticity implies that the estimated λ_{ij} are smaller in levels and less dispersed, as evidenced by comparing Tables 3 and D.1. The key pattern is still preserved in the case with higher κ: in 2015, it is significantly harder to move from the rural regions to the large cities ($\lambda_{21} = 8.06$) than to the smaller ones ($\lambda_{31} = 4.50$). The impacts of the alternative urbanization policies are qualitatively similar but quantitatively larger. Adopting the λ^* policy leads to 3.0%, and the "low λ" policy, a 18.2% increase in national welfare. These numbers are to be compared with the 2.6% and 15.6% welfare gains in the baseline. The welfare gains are higher here because the migration flows are more sensitive to the changes in λ_{ij} in a world with a high elasticity.

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	1.00	0.48	0.43
MUR (d)	8.06	1.00	2.42
OUR (d)	4.50	2.35	1.00
f_e	10.67		

(a) Higher Migration Elasticity, $\kappa = 3.3$

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	1.00	0.70	0.49
MUR (d)	21.88	1.00	5.15
OUR (d)	7.79	6.53	1.00
f_e	10.75		

(b) Higher Elasticity of Substitution, $\varepsilon = 10.0$

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	1.00	5.16	2.36
MUR (d)	2.97	1.00	3.39
OUR (d)	1.61	9.93	1.00
f_e	10.16		

(c) Higher Expenditure Share of Land Consumption, $1 - \alpha - \gamma = 0.25$

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	1.00	1.06	0.64
MUR (d)	14.33	1.00	4.48
OUR (d)	5.92	7.69	1.00
f_e	-		

(d) Fixed Entry

Table D.1: Robustness Checks, the Re-Calibrated Parameters: f_e and $\{\lambda_{ij}\}$

Note: This table reports the jointly calibrated parameters in the robustness checks. The other parameters are the same as in the 2015 baseline model.

D.2 Higher Elasticity of Substitution

In the baseline model, we jointly calibrate ε and θ to match a trade elasticity of 4 and a tail-index of firm-size distribution of 1.076. The resulting $\varepsilon = 4.717$ implies an average markup of 27%. In the robustness check, we increase the elasticity of substitution to $\varepsilon = 10.0$ so the market structure is closer to perfect competition with a markup of 11% while the trade elasticity, θ, is reset to 9.684 to match the tail-index of 1.076. The new values of ε and θ remain in the ballpark of the estimates from the gravity-equation literature.

With a lower markup, the real income level in all cases improves substantially, as lower market power increases the firms' equilibrium output. In this framework in which there is a differentiated sector with positive markups and a rural sector with zero markups, the equilibrium allocation is always sub-optimal as the allocation of labor to urban regions is less than optimal. A reduction in market power in the differentiated sector reduces this allocative inefficiency and implies a larger rural-urban migration.[2] However, to keep migration flows as the observed ones, the re-calibrated rural-to-urban migration frictions would be higher than the baseline ones. This is apparent from

[2]For the economics underlying allocative inefficiency due to variable markups, see, for example, Holmes et al. (2014) and Arkolakis et al. (2019).

	China	Rural	MUR	OUR	Rural ODC	Urban ODC	Rural ROW	Urban ROW
	Baseline 2015, $\kappa = 3.3$							
Population	2.1847	0.6624	0.5210	1.0013	1.4761	3.0510	0.0322	1.3668
Welfare	0.2084	0.0610	0.2824	0.2347	0.2748	2.1754	6.8823	10.2875
	$\lambda_{21} = \lambda_{31} = \lambda^*$, $\kappa = 3.3$							
Population	2.1847	0.6630	0.6473	0.8744	1.4761	3.0510	0.0322	1.3668
Welfare	0.2146	0.0619	0.2901	0.2375	0.2731	2.1623	6.8726	10.2083
	Low λ, $\kappa = 3.3$							
Population	2.1847	0.5669	0.7269	0.8908	1.4761	3.0510	0.0322	1.3668
Welfare	0.2463	0.0697	0.3137	0.2554	0.2748	2.1761	7.2709	10.2712

(a) Higher Migration Elasticity

	China	Rural	MUR	OUR	Rural ODC	Urban ODC	Rural ROW	Urban ROW
	Baseline 2015, $\varepsilon = 10.0$							
Population	2.1847	0.6630	0.5209	1.0008	1.4761	3.0510	0.0322	1.3668
Welfare	0.6645	0.1326	0.8926	0.6995	0.3537	2.7049	8.6927	11.0932
	$\lambda_{21} = \lambda_{31} = \lambda^*$, $\varepsilon = 10.0$							
Population	2.1847	0.6655	0.6241	0.8951	1.4761	3.0510	0.0322	1.3668
Welfare	0.6767	0.1347	0.9023	0.7037	0.3536	2.7040	8.7577	11.0844
	Low λ, $\varepsilon = 10.0$							
Population	2.1847	0.5853	0.6797	0.9197	1.4761	3.0510	0.0322	1.3668
Welfare	0.7368	0.1460	0.9217	0.7186	0.3537	2.7049	9.1854	11.0594

(b) Higher Elasticity of Substitution

Table D.2: Robustness Checks: Results I

Note: This table lists the key endogenous variables for all 7 regions across the baseline and the counter-factual simulations. The first column is the aggregate result for China.

comparing Tables 3g and D.1. Similarly, the urban-to-rural migration frictions are lower than the baseline ones.

The λ^* and the low-λ counterfactuals lead to the slightly lower welfare gains of 1.8% and 10.9%. In this framework, the elasticity of welfare to the allocation of labor is tied closely to the elasticity of substitution, which inversely reflects the love of variety. In the λ^* counter-factual, labor reallocation from the OUR to the MUR still brings welfare gains, but such gains become smaller when ϵ is higher because the new varieties that come with the inflow of population to the MUR are less valuable to consumers there. A similar logic applies to the low-λ counter-factual.

	China	Rural	MUR	OUR	Rural ODC	Urban ODC	Rural ROW	Urban ROW
	Baseline 2015, high expenditure share of land consumption							
Population	2.1847	0.6629	0.5209	1.0008	1.4761	3.0510	0.0322	1.3668
Welfare	0.3728	0.2975	0.2719	0.3237	1.4121	4.5223	16.6967	17.7111
	$\lambda_{21} = \lambda_{31} = \lambda^*$, high expenditure share of land consumption							
Population	2.1847	0.6636	0.5799	0.9412	1.4761	3.0510	0.0322	1.3668
Welfare	0.3746	0.3012	0.2758	0.3264	1.4111	4.5192	16.7040	17.6872
	Low λ, high expenditure share of land consumption							
Population	2.1847	0.6185	0.6005	0.9657	1.4761	3.0510	0.0322	1.3668
Welfare	0.3991	0.3174	0.2858	0.3385	1.4126	4.5238	16.9206	17.6325

(a) Higher Expenditure Share of Land Consumption

	China	Rural	MUR	OUR	Rural ODC	Urban ODC	Rural ROW	Urban ROW
	Baseline 2015, fixed entry							
Population	2.1847	0.6600	0.5168	1.0078	1.4761	3.0510	0.0322	1.3668
Welfare	0.3319	0.0889	0.3890	0.3586	0.4179	3.2927	10.8556	15.8012
	$\lambda_{21} = \lambda_{31} = \lambda^*$, fixed entry							
Population	2.1847	0.6589	0.5849	0.9409	1.4761	3.0510	0.0322	1.3668
Welfare	0.3312	0.0892	0.3734	0.3682	0.4177	3.2913	10.8619	15.7932
	Low λ, fixed entry							
Population	2.1847	0.5968	0.6178	0.9700	1.4761	3.0510	0.0322	1.3668
Welfare	0.3478	0.0935	0.3717	0.3697	0.4173	3.2885	11.0922	15.7584

(b) Fixed Entry

Table D.3: Robustness Checks: Results II

Note: This table lists the key endogenous variables for all 7 regions across the baseline and the counter-factual simulations. The first column is the aggregate result for China.

This pattern is similar to Ma and Tang (2020), who also document declining gains from migration as the elasticity of substitution increases.

D.3 Higher Expenditure Share of Land Consumption

In the baseline model, the expenditure share of land consumption, $1 - \alpha - \gamma$, is set to 0.0625 following Davis and Ortalo-Magné (2011) and Combes et al. (2019), as explained in Section 4.2.1. As this paper does not explicitly model housing structure, which is treated as part of the differen-

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.622	0.075	0.181
MUR (d)	0.254	0.898	0.082
OUR (d)	0.124	0.026	0.737

(a) λ^*, High Migration Elasticity

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.447	0.086	0.203
MUR (d)	0.373	0.889	0.081
OUR (d)	0.180	0.026	0.716

(b) Low λ, High Migration Elasticity

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.622	0.079	0.182
MUR (d)	0.229	0.894	0.079
OUR (d)	0.149	0.027	0.739

(c) λ^*, High Elasticity of Substitution

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.483	0.086	0.197
MUR (d)	0.314	0.887	0.077
OUR (d)	0.204	0.027	0.726

(d) Low λ, High Elasticity of Substitution

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.622	0.078	0.180
MUR (d)	0.166	0.895	0.079
OUR (d)	0.212	0.027	0.740

(e) λ^*, High Expenditure Share of Land Consumption

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.551	0.080	0.184
MUR (d)	0.196	0.893	0.079
OUR (d)	0.252	0.027	0.737

(f) Low λ, High Expenditure Share of Land Consumption

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.607	0.059	0.177
MUR (d)	0.111	0.920	0.105
OUR (d)	0.282	0.020	0.718

(g) Low growth of $\bar{R}_{\text{MUR}}$, High Expenditure Share of Land Consumption

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.619	0.073	0.179
MUR (d)	0.090	0.901	0.085
OUR (d)	0.291	0.026	0.737

(h) Low growth of $\bar{R}_{\text{MUR}}$, Baseline

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.622	0.085	0.173
MUR (d)	0.193	0.885	0.069
OUR (d)	0.184	0.030	0.758

(i) λ^*, Fixed Entry

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.515	0.092	0.184
MUR (d)	0.246	0.878	0.067
OUR (d)	0.238	0.031	0.749

(j) Low λ, Fixed Entry

Table D.4: Robustness Checks: Migration Probability Matrices

Note: This table presents the migration probability matrix within China in various robustness checks. An element at the i-th row and the j-th column indicates the probability of an individual originating from j and moving to i. Each column sums to 1.

tiated goods, we consider this number an appropriate one to use. The finding that households are not entirely responsive to land prices may be likely because this expenditure share is low. Thus, we conduct a robustness check by experimenting with a higher expenditure share at 0.25, which can be considered on the high end of possible values of the expenditure share of land consumption.

With a higher weight on land consumption, the MUR and OUR's high land prices deter rural migrants. As a result, the estimated rural-to-urban migration frictions drastically decline to $\lambda_{31} = 2.97$, and $\lambda_{21} = 1.61$. Nevertheless, the migration barriers into the MUR are still higher than that into the OUR, reflecting the discriminatory urbanization policy. We first repeat the "low growth of $\bar{R}_{\text{MUR}}$" counterfactual exercise. Naturally, we find that rural emigrants are more responsive to the changes in land prices than the baseline model. With a lower growth rate of $\bar{R}_{\text{MUR}}$, the rural-to-MUR migration probability increased to 11.1% as shown in Table D.4. In comparison, in the baseline model, the same probability is only 9.0%.

Our main results are robust to this alternative parameterization. The λ^* and low-λ policies still divert a significant proportion of the rural emigrants towards the MUR (see Table D.4[e,f]). However, the magnitudes are smaller than the baseline case (see Table 5[b,c]). Importantly, the alternative urbanization policies are still more effective in re-directing the population flows than depressing the growth rates of $\bar{R}_{\text{MUR}}$. This is evident from comparing Panels (e) and (f) with Panel (g) in Table D.4.

Lastly, the welfare impacts of the alternative urbanization policies become relatively mild. As shown in Table D.3, the improvements of national welfare under the λ^* and low-λ policies are reduced to 0.5% and 7.1%, respectively. This is because the alternative policies attract a smaller fraction of the rural emigrants into the more productive MUR if the individuals care more about land consumption.

D.4 Fixed Entry

The firm-entry margin is instrumental to the punchline result that a more uniform or laissez-faire migration policy improves national welfare. To highlight the role of firm entry in our model, we shut down the firms' entry-and-exit channel. In the baseline model, I_j potential firms pay the entry fee f_e, and we need to solve for I_j in the general equilibrium. For the "fixed entry" model, we assume that $\bar{I}_j$ is exogenously given at a level that will be specified later, and the entry cost f_e is assumed to be zero. Without the firm-entry margin, the aggregate profit becomes positive (instead of zero in the baseline model). Regional aggregate profits are evenly rebated to individuals in that region. In the new version, we provide the details of solving the model in this new setup in

Appendix B.7.

To study the effect of entry on migration, we compute $\bar{I}_j$'s used in all fixed-entry exercises from the pre-migration equilibrium, i.e., the equilibrium under the initial population and before people move. In practice, this is computed from the baseline model with free entry and prohibitive migration costs ($\lambda_{ij} = \infty, i \neq j$). Then, any equilibrium with the fixed $\bar{I}_j$'s is a world in which the number of entrants no longer responds to population flows.

The estimated λ_{ij}'s reported in Table D.1 are similar to the baseline case, in which the barriers to the MUR are significantly higher than those to the OUR. Similarly, the λ^* exercise also leads to a pronounced shift of the rural emigrants towards the MUR, although the magnitude is slightly smaller. See Table D.4(i) and compare it with Table 5(b).

Nevertheless, the welfare impacts of alternative policies are drastically different. In the fixed-entry model, the MUR suffers lower welfare when more rural migrants flow into the large cities in the λ^* exercise. As a result, national welfare drops. This is in stark contrast with the result in the baseline model in which all regional and national welfare increases. As this result in the baseline model is the paper's punchline, this exercise under the fixed-entry model highlights the importance of the entry margin. Under the fixed mass of firms, inflows of people no longer increase the number of varieties in the destination market; instead, they only push down the factor prices, push up the land prices, and eventually reduce local welfare. The importance of firm entry in the context of migration is already highlighted in Ma and Tang (2020), and this robustness check resonates with their finding.

The result in the low-λ policy is similar in terms of the directions of changes in regional welfare, but the national welfare still improves as the overall migration frictions are lowered. All of the welfare results of the two alternative migration policies are reported in Table D.3(b).

E Additional Results

E.1 Reverting Productivity and Amenity

In Section 4.3, we note that the evolution of productivity and amenity do not explain the observed pattern based on the estimated parameters. In this appendix, we conduct two counter-factual analyses to evaluate the impacts of these two channels on the migration flows. In these two exercises, we revert the productivity and amenity estimates to the 2005 levels respectively while keeping all the other parameters the same as in the baseline model in 2015. Table E.1 reports the migration probabilities under these two counter-factual exercises.

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.666	0.093	0.210
MUR (d)	0.074	0.880	0.076
OUR (d)	0.260	0.027	0.715

(a) Reverting μ_i

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.606	0.090	0.165
MUR (d)	0.069	0.875	0.062
OUR (d)	0.325	0.036	0.773

(b) Reverting ϕ_i

Table E.1: Matrices of Migration Probability, Reverting Productivity and Amenity

Note: This table presents the matrices of migration probability. An element at the i-th row and the j-th column indicates the probability of an individual originating from j and moving to i. Each column sums to 1. The data source is the *One-Percent Population Survey* in the respective years, and an "origin" is defined as the place of hukou registration.

Neither productivity nor amenity explain the observed pattern of migration probability. Under the productivity in 2005, rural migrants are still 2.5 time more likely to move to the OUR, the same as in the baseline. This is expected as the relative productivity between the MUR and OUR changes little between the two years. Similarly, the evolution of amenities does not explain the migration pattern in 2015. Reverting the amenity leads to an even stronger preference for the OUR. This result is, again, expected as the amenities of the OUR are stronger than those of the MUR in 2005.

E.2 The Role of the Fixed Exporting Barriers

In the Melitz framework, the decision of where to sell goods is captured by the fixed exporting costs, f_{ij}. Such barriers shape firms' decisions as to which markets to sell. However, we find that market selection does not interact with the urbanization policy.

To highlight the irrelevance of market selection, we simulate a counterfactual in which f_{ij}'s are reduced to half the values in the baseline. With lower barriers to export, more firms engage in international trade. Under both the baseline and the alternative urbanization policies, the fraction of exporting firms is higher than the baseline quantification. Similarly, the welfare also increases with the lowered f_{ij}'s due to the gains from trade.

	China	Rural	MUR	OUR	Rural ODC	Urban ODC	Rural ROW	Urban ROW
	Baseline 2015, low f_{ij}							
Population	2.1847	0.6627	0.5210	1.0010	1.4761	3.0510	0.0322	1.3668
Welfare	0.2268	0.0611	0.2838	0.2359	0.2748	2.1756	6.8795	10.2768
Operating Firms	2.4575	-	1.2328	1.2247	-	189.5376	-	6187.4230
Exporting Firms	0.5727	-	0.2873	0.2854	-	15.0688	-	70.0431
	$\lambda_{21} = \lambda_{31} = \lambda^*$, low f_{ij}							
Population	2.1847	0.6628	0.6208	0.9010	1.4761	3.0510	0.0322	1.3668
Welfare	0.2325	0.0622	0.2916	0.2393	0.2747	2.1750	6.9116	10.2713
Operating Firms	2.2649	-	1.2983	0.9666	-	183.8977	-	6005.4568
Exporting Firms	0.5878	-	0.3370	0.2509	-	14.6224	-	67.9738
	Low λ, low f_{ij}							
Population	2.1847	0.5766	0.6764	0.9317	1.4761	3.0510	0.0322	1.3668
Welfare	0.2619	0.0685	0.3095	0.2533	0.2749	2.1762	7.2123	10.2623
Operating Firms	1.6435	-	0.9636	0.6799	-	118.5085	-	3882.7503
Exporting Firms	0.4137	-	0.2426	0.1712	-	9.4340	-	43.8969

(a) Welfare

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.622	0.077	0.180
MUR (d)	0.221	0.896	0.080
OUR (d)	0.156	0.027	0.740

(b) Migration Probability, λ^*

	Rural (o)	MUR (o)	OUR (o)
Rural (d)	0.484	0.082	0.190
MUR (d)	0.303	0.892	0.079
OUR (d)	0.213	0.027	0.731

(c) Migration Probability, low λ

Table E.2: Lowering the Fixed Costs of Exporting

Note: This table shows the welfare impacts of reducing the fixed costs of exporting.

However, the migration patterns and the welfare impacts of the alternative urbanization policies are remarkably similar to those under the baseline quantification. The lack of interaction between exporting behavior and migration is due to the absence of intranational geography in the paper, without which the MUR and OUR have equal access to the world market. As a result, variations in the fixed exporting costs affect both regions equally, leaving little room to interact with the urbanization policies.

Incorporating the intranational geography will only strengthen our main results. This is because most of the megacities in the MUR region are in the coastal areas, which enjoy lower exporting costs (both fixed and variable ones). Thus, the effects of a more uniform or laissez-faire urbanization policy will be even larger.

References

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare, "The elusive pro-competitive effects of trade," *The Review of Economic Studies*, 2019, *86* (1), 46–80.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, "The costs of agglomeration: House and land prices in French cities," *The Review of Economic Studies*, 2019, *86* (4), 1556–1589.

Davis, Morris A and François Ortalo-Magné, "Household expenditures, wages, rents," *Review of Economic Dynamics*, 2011, *14* (2), 248–261.

Donaldson, Dave and Richard Hornbeck, "Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 02 2016, *131* (2), 799–858.

Holmes, Thomas J, Wen-Tai Hsu, and Sanghoon Lee, "Allocative efficiency, mark-ups, and the welfare gains from trade," *Journal of International Economics*, 2014, *94* (2), 195–206.

Ma, Lin and Yang Tang, "Geography, trade, and internal migration in China," *Journal of Urban Economics*, 2020, *115*, 103181. Cities in China.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg, "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, December 2018, *108* (12), 3855–90.