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

Wen-Tai Hsu† Lin Ma‡

March 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 and by the coastal-inland divide. We find a striking contrast in the migration patterns between years 2005 and 2015; rural people tended to move more to the coastal urban region in 2005, but more to the inland urban region in 2015. We calibrate a spatial quantitative model to the world economy in both years with China being divided into the rural, coastal urban, and inland urban regions. We find that alternative urbanization policies that are not differential and that are more laissez-faire would 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.
†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, of course, aware of 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 on the hukou system have therefore been slow and somewhat difficult, and thus the system has remained until today. While more background information on the hukou system will be given in Section 2, our key observation is that there is a prevalent thinking in China that the largest cities are too large and hence hukou restrictions in these cities are still direly needed to “control” the growth of these cities, despite the fact that the government and the general public also recognize the benefits of rural-urban migration for rural people and the overall economy.

It is, then, not surprising to find that the reforms on the hukou system have been differential in city size. The reforms greatly relaxed the restrictions to obtain a hukou in small- and medium-sized cities beginning around 2010, and the restrictions to obtain a hukou in these cities were totally abandoned around 2014–2015. Meanwhile, there remain conditions for migrants to obtain hukous in large cities, and it is still very difficult for the few largest cities. Government documents have shown that the thoughts of these differential reforms began to take shape around 2006, and the reforms gradually took place during 2006–2015. 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 the years 2005 and 2015. We find that even though there was more rural-urban migration in 2015 than in 2005,
the movement was more toward coastal cities in 2005, whereas the tendency shifted drastically toward inland cities in 2015. This finding is consistent with the above-mentioned reforms. We then propose a story with a global perspective that differential reforms may be harmful to the Chinese economy compared with the cases in which reforms are not differential. We calibrate a quantitative spatial model to the world economy in which China is divided into three regions – rural, coastal urban, and inland urban. We then conduct quantitative analyses to evaluate these differential reforms by comparing 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 coastal urban areas are most productive in China, as they contain 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-Tianjing agglomeration, and many other prosperous cities. City sizes in this region is typically much larger than those outside this region; China’s four largest cities (Shanghai, Shenzhen, Beijing, and Guangzhou) are also in this region. The differential reforms thus did not relax much the restrictions on obtaining hukous in coastal cities; they mainly helped relax the restrictions on obtaining hukous in inland cities. Second, out of various concerns over regional inequality, there have been various ongoing programs targeting infrastructure building in the inland area, e.g., the Western Development Program (Xi Bu Da Kai Fa). Such programs may also change the relative migration friction between rural-inland-urban and rural-coastal-urban migration, and justify our specific partition of China. Third, we opt not to go for a many-region model within China, as we seek to sharpen the contrast between coastal and inland urban regions and to allow room to incorporate other countries in the world.

Our story is briefly as follows. First, we do not believe there is an a priori reason to think that China’s largest cities are too large; evaluating optimal city size is a delicate and complicated issue, and more scientific studies along the lines such as Au and Henderson (2006) will be needed. However, policies induced by such prevalent thinking may be harmful for China’s economic development because when resources/population are reallocated away from the most productive area, i.e., the coastal urban region, due to the differential in migration friction, the aggregate productivity
may be lowered. In various policy narratives, the government would like to promote industrialization in the inland urban region. In other words, the government has a preference of moving firms to inland urban areas over moving people to the coastal urban areas. This would result in coastal urban wages rising faster than in the case without such preference. An important neglect of such policy narratives is that firms need not move to China’s inland urban areas if they think 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 inland urban areas, the entry of firms there would also increase. Therefore, relative productivity and wages among ODC and China’s coastal and inland urban areas will be 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 and rural 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 or not to migrate to a different location based on their idiosyncratic locational preferences, migration friction, and consumption utility derived from location-specific 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.

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. The key calibration object is the migration costs within China; for this purpose, we utilize the information of migration flows revealed in the Intercensal Population Sample Survey of One-Percent (henceforth One-Percent Population Survey for short) in 2005 and 2015. In these two surveys, surveyed individuals were asked their hukou registra-
tion 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. Another important calibration target is the region-specific productivity. Location choices are only partly determined by migration costs; they also depend on relative productivity across locations, which, in turn, determines relative wages and consumption prices along with trade costs in the general equilibrium.

The calibrated migration costs for rural-coastal-urban and rural-inland-urban migration are similar in 2005; nonetheless, while both migration costs drop in 2015, the drop for the rural-inland-urban is much sharper than that for rural-coastal-urban, rationalizing the above-mentioned migration patterns. A potential threat to our theory is that the gap of China’s inland urban region’s productivity from the coastal urban region’s may have risen over the 2005–2015 period. In this case, “regional convergence” in productivity would cause rural-urban migration to shift toward inland urban as observed in the data. However, our calibrated results show that the relative productivity between the two regions barely changed over this 10-year span.

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-coastal-urban and rural-inland-urban migration to be the same at the level where the 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 the rural-coastal-urban migration being larger than the rural-inland-urban migration. This is natural because the coastal urban region is more productive and migration costs are made the same. The resulting reallocation of labor improves national welfare by 2.2%; this magnitude is substantial considering that this involves only 2.9% of the total population reallocated from the inland urban region to the coastal urban one.

In the second counter-factual exercise, we consider a more liberalized version of the urbanization policy in which the rural-coastal-urban migration cost is set the same as the rural-inland-urban 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 coastal urban region compared with the first counter-factual. The resulting improvement in national welfare is larger than the first counter-factual at 6.9%.
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 4.4% and 12.7% 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 more laissez-faire urbanization policies, China may 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., the urbanization policy that is 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-city 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. 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 we focus on the fact that reforms on the hukou system, particularly those after 2010, are differential – discriminating large and coastal cities –, 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 of 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 quantitative analyses 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.

### 2.2 Migration Patterns

In this subsection, we examine migration patterns in 2005 and 2015 using the information of migration flows from the *One-Percent Population Surveys* in 2005 and 2015. This population intercensal survey is part of China’s census program. The survey attempted 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
Table 1: Matrices of Migration Probability

<table>
<thead>
<tr>
<th></th>
<th>Rural (d)</th>
<th>Coastal (d)</th>
<th>Inland (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural (o)</td>
<td>0.945</td>
<td>0.054</td>
<td>0.064</td>
</tr>
<tr>
<td>Coastal (d)</td>
<td>0.036</td>
<td>0.943</td>
<td>0.018</td>
</tr>
<tr>
<td>Inland (d)</td>
<td>0.020</td>
<td>0.004</td>
<td>0.917</td>
</tr>
</tbody>
</table>

(a) Migration Probability, 2005

<table>
<thead>
<tr>
<th></th>
<th>Rural (d)</th>
<th>Coastal (d)</th>
<th>Inland (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural (o)</td>
<td>0.624</td>
<td>0.133</td>
<td>0.212</td>
</tr>
<tr>
<td>Coastal (d)</td>
<td>0.149</td>
<td>0.851</td>
<td>0.065</td>
</tr>
<tr>
<td>Inland (d)</td>
<td>0.227</td>
<td>0.016</td>
<td>0.723</td>
</tr>
</tbody>
</table>

(b) Migration Probability, 2015

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.

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 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 $j$.

To sharpen both the empirical examination and quantitative analysis, we divide China into three regions: rural, coastal urban, and inland urban. We include all the urban areas in the coastal provinces as the “coastal urban” region. The other urban areas in China are then grouped in the “inland urban” region. The rest of China is then the “rural” region.

Table I shows the matrices of migration probability in 2005 and 2015. In 2005, 3.6% of rural individuals migrated to the coastal urban region, whereas 2.0% of them migrated to the inland urban region. In 2015, these numbers rose sharply to 14.9% and 22.7%, 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 the migration probability to the coastal urban region was 1.8 times larger than that to the inland urban region in 2005, it became only 65.6% of the probability to the inland region 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

---

1The coastal provinces include Guangdong, Fujian, Zhejiang, Jiangsu, and Shandong. In addition, we also include the three direct-control municipalities (Zhi Xia Shi) on the east coast in the “coastal urban” region: Beijing, Shanghai, and Tianjin.
inland and 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 The Model

Our model follows the framework in Ma and Tang (2020b), which builds on Melitz (2003) and Tombe and Zhu (2019).

3.1 Basic Environment

The world consists of \(N\) 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 \(\bar{L}_c\) 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 the differentiated products as both the manufactured goods and the services. Both the 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.

---

2There 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.

3As 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.
3.2 Consumption

Individuals in each location $j$ derive utility from consumption and idiosyncratic locational preferences, and 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 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 and a CES composite of differentiated goods:

$$U_i = \left( y_i^A \right)^{\alpha} \left[ \sum_{j=1}^{J} \int_{k \in \Omega_{ij}} y_{ij}(k) \frac{\varepsilon - 1}{\varepsilon - 1} dk \right]^{\frac{\varepsilon - 1}{\varepsilon - 1}}, \quad 0 < \alpha < 1,$$

(1)

where $y_{ij}(k)$ 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 the agriculture product, $\alpha$ captures the expenditure share of the agricultural good, and $\Omega_{ij}$ denotes the set of varieties from location $j$ available for purchase in location $i$ and is endogenously determined.

The set of varieties consumed in location $i$, $\Omega_i \equiv \bigcup_j \Omega_{ij}$, depends directly on firm entry and exit decisions in $j$, 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 The 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 \( \tau_{ij} \) 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 \( b_i(k) \):

\[
q_i(k) = \frac{1}{a(k)} b_i(k),
\]

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 \( i \):

\[
b_i(k) = [\ell_i(k)]^\beta \left[ \left( \sum_{j=1}^{J} \int_{k' \in \Omega_{ij}} y_{ij}(k'; k) \frac{x_k}{x} \, dk' \right)^{1-\beta} \right],
\]

where \( \ell_i(k) \) 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 \( \beta \) is the relative weight of labor in the production function.

Firms are heterogeneous in productivity, as \( a(k) \), the input bundle requirements 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\(^4\)

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

where \( \theta \) is the tail index and \( \mu_j \) is the parameter that reflects the average productivity in location \( j \). A higher \( \mu_j \) implies that the average draw of \( a \) is lower in \( j \), and \( 1/\mu_j \) defines the maximum of \( a \).

---

\(^4\)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).
The cumulative distribution function (CDF) of $a$ is therefore

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

**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.

### 3.4 The Agriculture Sector

The production in the agriculture sector is linear in labor input, subject to a productivity parameter $\mu_j^A$:

$$q_j^A = \mu_j^A \cdot \ell_j^A. \quad (4)$$

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 rural wage rate:

$$P_{ij}^A = \frac{\tau_{ij}^A w_j}{\mu_j^A}. \quad (5)$$

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

$$P_i^A = \min_{j=1,\ldots,J} \{ P_{ij}^A \}. \quad (6)$$

### 3.5 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 is given by

\[ U_i = \left[ \frac{\alpha w_i}{P_i^A} \right]^\alpha \left[ \frac{(1 - \alpha)w_i}{P_i} \right]^{(1-\alpha)}, \]

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

\[ P_i = \left[ \sum_{j=1}^{J} \int_{\Omega_{ij}} (p_{ij}(k))^{1-\varepsilon} dk \right]^{\frac{1}{1-\varepsilon}}. \quad (7) \]

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 \( \iota_i \) is i.i.d across locations and individuals. We assume that \( \iota_i \) follows a Fréchet distribution with CDF

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

where \( \kappa \) is the shape parameter that controls the heterogeneity of these locational-preference shocks; the smaller the \( \kappa \), the larger the heterogeneity. Such locational-preference specification has been well understood as a dispersion force, which is stronger if the heterogeneity is larger (\( \kappa \) 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 \( c \):

\[ \frac{U_i \cdot \iota_i}{\lambda_{ij}} \geq \frac{U_{i'} \cdot \iota_{i'}}{\lambda_{i'j}}, \quad \forall i' \in J_c. \]
3.6 Equilibrium

Let $p_{ij}(.)$ and $q_{ij}(.)$ denote the profiles of price and total quantity sold from $j$ to $i$ across $\Omega_{ij}$, respectively. Let $L_j$ 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 readers to Appendix B for more details.

**Definition:** An equilibrium consists a tuple of prices $\{w_j, p_{ij}(.), p^A_j\}_{i,j}$ and a tuple of quantities $\{L_i, I_i, q_{ij}(.), q^A_i, y_{ij}(.), y^A_i\}_{i,j}$ such that the following conditions hold:

(a) Individuals maximize their utility by choosing locations and 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) The differentiated goods market clears such that the aggregate expenditure on the differentiated goods in location $i$ equals the final consumption $(1 - \alpha)w_iL_i$ and intermediate goods use $(1 - \beta)X_i$: $X_i = (1 - \alpha)w_iL_i + (1 - \beta)X_i$.

(f) Labor market clearing for each country $c$: $\sum_{j \in J_c} L_j = \bar{L}_c$.

3.7 Analytical Solutions

We sketch the analytical solution in this subsection, and refer the readers to Appendix B for the details.
3.7.1 The 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 \( i \):

\[
q_{ij}(k) = \frac{X_i}{(P_i)^{1-\varepsilon}} [p_{ij}(k)]^{-\varepsilon}.
\] (8)

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 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 (8), and \( \chi_j \) is the cost of an input bundle in the differentiated sector at location \( j \), which itself is the solution of a cost minimization problem:

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

Standard procedure yields the constant-markup pricing:

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

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 \( \varepsilon \) is greater than 1. The variable profit is also higher.
for firms with lower $a(k)$ as $\pi_{ij}$ is proportional to $(a(k))^{1-\varepsilon}$:

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

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 the 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, $f_{ij}$: $\pi_{ij}(a) \geq f_{ij}$. 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 $a_{ij}, i = 1, 2, \ldots, J$ 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 $i$:

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

where $1 (a(k) < a_{ij})$ 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.

**The 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

$$\Pi_j \equiv E [\Pi_j(a(k))] = \int_{0}^{1/\mu_j} \Pi_j(a) dG_j(a).$$
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:

\[
\Pi_j = f_e \chi_j. \tag{9}
\]

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 \( j \):

\[
P_i = \left[ \sum_{j=1}^{J} \left( \frac{\varepsilon}{\varepsilon - 1} \tau_{ij} \chi_j \right)^{1-\varepsilon} I_j \int_{0}^{a_{ij}} a^{1-\varepsilon} dG_j(a) \right]^{\frac{1}{1-\varepsilon}},
\]

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.

### 3.7.2 Migration Decision

It is straightforward to show that conditional on \( \{U_i\}_{i \in J_c} \), the fraction of the population that migrates from \( j \) to \( i \) in country \( c \) is

\[
m_{ij} = \frac{(U_i)^{\kappa} (\lambda_{ij})^{-\kappa}}{\sum_{i' \in J_c} (U_{i'})^{\kappa} (\lambda_{i'j})^{-\kappa}}. \tag{10}
\]

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 \( \kappa \) is the migration elasticity with respect to friction. The larger the \( \kappa \), 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 64 countries in the sample, we group the countries with the average per capital GDP less then 2/3 of the USA into the “other developing countries”, and the rest as the ROW. Appendix C provides the details about the sample definition, and Table A.1 in the appendix lists all the countries in the sample. As mentioned in Section 2.2, China is divided into three regions: rural \( r = 1 \), the coastal urban \( r = 2 \), and the inland urban \( r = 3 \). The other two countries only contain one rural and one urban region each.

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 readers to Appendix C for more details. All parameters are summarized in Table 2.

4.1 Common Parameters

This group of parameters is common across the two years:

- The expenditure share on agriculture products, \( \alpha = 0.15 \). We back out this parameter using the household expenditure on agriculture products as a share of total consumption. The data source is the Input-Output Table from China in the year 2002.

- The labor share in differentiated products, \( \beta = 0.37 \). The data source is also the IO table in China. 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, \( \varepsilon \), and the Pareto tail index, \( \theta \). 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.\(^5\)

- 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 U.S. and Hsieh and Moretti (2019) set it to 2.0 based on a similar extreme-value distribution. Bryan 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 \( \kappa = 3.3 \), the estimates from the higher end of the spectrum, as a robustness check in Section 5.3. As will be seen in Section 5.3 all the main welfare results are strengthened in the 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.

### 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.

---

\(^5\)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.
4.2.1 Initial Population, Trade Cost, and Productivity

**Initial Population** To construct the initial population distribution for the 2005 model and 2015 one, 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 with the percentage of workforce employed in agriculture from the WDI delivers the rural population in each country. The total urban population is simply the difference between the total and the rural population. Within China, the total urban population must be further divided between the coastal and the inland cities. 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, $\tau_{ij}$, 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 the iceberg friction, and $f_{ij} = f_{jj}$ so that the firms originating from region $j$ face the same fixed costs in selling locally and to another region $i$ in the same country.

We estimate the variable trade costs for agricultural goods as proportional to $\tau_{ij}$, 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

---

6 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 the counter-factual simulations.
the estimate of $\bar{\tau}$ and arrive at $\bar{\tau} = 2.14$.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Source</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.15</td>
<td>Input-Output Table, 2002</td>
<td>Expenditure share in agricultural goods</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.37</td>
<td>Input-Output Table, 2002</td>
<td>Labor share in differentiated goods production</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>4.717</td>
<td>Firm size distribution in China</td>
<td>Elasticity of substitution</td>
</tr>
<tr>
<td>$\theta$</td>
<td>4.0</td>
<td>Trade elasticity</td>
<td>Pareto tail index in productivity distribution</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>1.63</td>
<td>Migration elasticity</td>
<td>Shape parameter in location preference</td>
</tr>
</tbody>
</table>

(a) Common Parameters

<table>
<thead>
<tr>
<th></th>
<th>Rural China</th>
<th>Coastal Urban</th>
<th>Inland Urban</th>
<th>Rural ODC</th>
<th>Urban ODC</th>
<th>Rural ROW</th>
<th>Urban ROW</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_i$, 2005</td>
<td>1.00</td>
<td>0.35</td>
<td>0.71</td>
<td>1.62</td>
<td>2.41</td>
<td>0.04</td>
<td>1.28</td>
</tr>
<tr>
<td>$L_i$, 2015</td>
<td>0.71</td>
<td>0.45</td>
<td>1.03</td>
<td>1.48</td>
<td>3.05</td>
<td>0.03</td>
<td>1.38</td>
</tr>
<tr>
<td>$\mu_i$, 2005</td>
<td>1.00</td>
<td>1.00</td>
<td>0.85</td>
<td>1.20</td>
<td>1.15</td>
<td>41.19</td>
<td>2.33</td>
</tr>
<tr>
<td>$\mu_i$, 2015</td>
<td>1.94</td>
<td>1.29</td>
<td>1.11</td>
<td>1.65</td>
<td>1.52</td>
<td>54.03</td>
<td>2.63</td>
</tr>
</tbody>
</table>

(b) Productivity and Initial Population

<table>
<thead>
<tr>
<th></th>
<th>China (o)</th>
<th>ODC (o)</th>
<th>ROW (o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>China (d)</td>
<td>1.00</td>
<td>2.59</td>
<td>2.55</td>
</tr>
<tr>
<td>ODC (d)</td>
<td>2.59</td>
<td>1.00</td>
<td>2.07</td>
</tr>
<tr>
<td>ROW (d)</td>
<td>2.55</td>
<td>2.07</td>
<td>1.00</td>
</tr>
<tr>
<td>(c) $\tau_{ij}$, 2005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China (d)</td>
<td>1.00</td>
<td>2.50</td>
<td>1.79</td>
</tr>
<tr>
<td>ODC (d)</td>
<td>2.50</td>
<td>1.50</td>
<td>2.30</td>
</tr>
<tr>
<td>ROW (d)</td>
<td>1.79</td>
<td>2.30</td>
<td>0.79</td>
</tr>
<tr>
<td>(d) $\tau_{ij}$, 2015</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China (d)</td>
<td>1.00</td>
<td>1.96</td>
<td>1.32</td>
</tr>
<tr>
<td>ODC (d)</td>
<td>1.96</td>
<td>0.96</td>
<td>1.28</td>
</tr>
<tr>
<td>ROW (d)</td>
<td>1.32</td>
<td>1.28</td>
<td>0.32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Rural (o)</th>
<th>Coastal (o)</th>
<th>Inland (o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural (d)</td>
<td>1.00</td>
<td>4.28</td>
<td>5.25</td>
</tr>
<tr>
<td>Coastal (d)</td>
<td>10.17</td>
<td>1.00</td>
<td>15.35</td>
</tr>
<tr>
<td>Inland (d)</td>
<td>10.42</td>
<td>20.93</td>
<td>1.00</td>
</tr>
<tr>
<td>$f_e$</td>
<td>11.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(g) Joint Calibration of $\lambda_{ij}$ and $f_e$, 2005</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Rural (o)</th>
<th>Coastal (o)</th>
<th>Inland (o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural (d)</td>
<td>1.00</td>
<td>2.34</td>
<td>2.21</td>
</tr>
<tr>
<td>Coastal (d)</td>
<td>3.21</td>
<td>1.00</td>
<td>6.12</td>
</tr>
<tr>
<td>Inland (d)</td>
<td>1.79</td>
<td>8.31</td>
<td>1.00</td>
</tr>
<tr>
<td>$f_e$</td>
<td>16.70</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(h) Joint Calibration $\lambda_{ij}$ and $f_e$, 2015

Table 2: Calibration Results

Note: This table summarizes the calibrated model parameters. Panel (a) presents the parameters that are common across the years, Panels (b)–(h) present the year-specific parameters.

**Productivity** 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 with estimating the productivity for each of the 279 cities (which corresponds 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 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 up 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 shall 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 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 coastal urban region 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 2b, we find that the relative productivity between the two urban regions in China barely moves between the two years. The inland productivity is 85.08% of the coastal productivity back in 2005, and the relative productivity only slightly increases to 85.37% 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.

The rural productivity is proxied by the per capita crop production data from the WDI. We normalize all the rural productivity so that the productivity in China is 1.0 in the year 2005.\(^7\)

\(^7\)In the quantification we have normalized both the productivity in coastal-urban and rural regions in China to 1.0.
4.2.2 Joint Calibration

The above quantification procedure leaves us with 7 parameters to jointly calibrate in each year. These parameters are the costs of entry, \( f_e \), and the six off-diagonal terms in the migration cost matrix in China, \( \lambda_{ij} \). The 7 parameters are jointly backed out by targeting 7 moment conditions in the model.

The cost of entry determines the number of entrants in the urban locations. We use this parameter to match the firms-to-population ratio in the coastal urban region 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 8.1 and 9.6 entering firms per thousand population.

The last six \( \lambda_{ij} \) parameters are pinned down by the migration probability matrix estimated from *One-Percent Population Survey* and presented in Table 1. The migration friction \( \lambda_{ij} \) are backed out by matching the model migration probability \( m_{ij} \) given in (10) to the observed migration probability \( \tilde{m}_{ij} \) in Table 1.

4.3 Changes in Migration friction

This subsection discusses the changes in migration friction in detail. The migration costs, \( \{\lambda_{ij}\} \), decline substantially between 2005 and 2015. The average magnitude drops by 64% 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 coastal urban regions drop by 68%, and to the inland urban areas, 83%. The changes in the other bilateral migration costs are lower and yet still sizable. The sharp decline in the calibrated \( \lambda_{ij} \) 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 individuals originating from the rural area choose to stay in the rural area, while 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 the inland cities. In 2005, the costs of moving from rural to
the two urban areas in China are roughly the same at $\lambda_{21} = 10.17$ and $\lambda_{31} = 10.42$. However, 10 years later, it is significantly easier to move to the inland cities ($\lambda_{31} = 1.79$) than to the coastal ones ($\lambda_{21} = 3.21$) from the rural area. The migration probability from the data also reveals the same pattern: the rural population is 80% more likely to move to coastal cities than to inland cities in 2005; fast-forward to 2015, and the pattern reverses and the rural population is 52% more likely to migrate to the inland cities. As our productivity measures have shown, the reversal in migration flows is unlikely to be driven by productivity changes, as we see little evidence of regional convergence in productivity. Therefore, the contrast between the migration patterns in these two periods are mainly explained by the changes in migration friction, which, in turn, is consistent with the differential reforms on the hukou system as discussed in Section 2.

## 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 the basic migration pattern that is consistent with the urbanization policy pivoting towards small-and-medium-sized and inland cities. In the model, the emphasis towards smaller, inland cities is reflected in the $\lambda$ matrix: while the migration costs from the rural region to the inland urban region dropped from 10.42 in 2005 to 1.79 in 2015, the rural-coastal-urban migration costs declined more mildly from 10.17 to 3.21. As a result, by 2015, out of the 37.6% of the rural population that emigrates, the majority, 22.7%, goes to the inland urban area and only 14.9% chooses to move to the coastal area.

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 “$\lambda^*$ counter-factual”, we eliminate the pivot towards inland cities by equalizing the rural-urban migration costs across the two urban

---

8From Table 1, $3.6/2.0 - 1 = 80\%$, and $22.7/14.9 - 1 \approx 52\%$
Table 3: Impact of Urbanization Policies

Note: The 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.

regions, e.g., setting $\lambda_{31} = \lambda_{21} = \lambda^*$. We pick the value of $\lambda^*$ so that the same 37.6% 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 $\lambda^* = 2.43$; as a result, the inbound friction towards the coastal regions is relaxed, while the friction into the inland regions is tightened. In the second counter-factual analysis, referred to as the “low $\lambda$ counter-factual”, we equalize the two rural-urban migration friction to the lower value of the two, so that $\lambda_{31} = \lambda_{21} = \min\{\lambda_{31}, \lambda_{21}\} = 1.79$. The rationale behind this exercise is to extend the more liberal migration policy to the rural-coastal-urban migration as well. Note that as this exercise is an overall relaxation of migration friction, we expect a higher volume of migration flow 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 the third panels of Table 3, where the first panel presents the baseline in 2015 for comparison.
5.1.1 $\lambda^*$ counter-factual

We first focus on the $\lambda^*$ counter-factual in which the same number of people move out of the rural region as compared to the baseline.

<table>
<thead>
<tr>
<th></th>
<th>Rural (o)</th>
<th>Coastal (o)</th>
<th>Inland (o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural (d)</td>
<td>0.624</td>
<td>0.131</td>
<td>0.212</td>
</tr>
<tr>
<td>Coastal (d)</td>
<td>0.239</td>
<td>0.853</td>
<td>0.065</td>
</tr>
<tr>
<td>Inland (d)</td>
<td>0.138</td>
<td>0.016</td>
<td>0.723</td>
</tr>
</tbody>
</table>

(a) $\lambda^*$

<table>
<thead>
<tr>
<th></th>
<th>Rural (o)</th>
<th>Coastal (o)</th>
<th>Inland (o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural (d)</td>
<td>0.625</td>
<td>0.134</td>
<td>0.214</td>
</tr>
<tr>
<td>Coastal (d)</td>
<td>0.149</td>
<td>0.850</td>
<td>0.064</td>
</tr>
<tr>
<td>Inland (d)</td>
<td>0.226</td>
<td>0.016</td>
<td>0.722</td>
</tr>
</tbody>
</table>

(c) Welfare-equivalent trade liberalization, $\lambda^*$

<table>
<thead>
<tr>
<th></th>
<th>Rural (o)</th>
<th>Coastal (o)</th>
<th>Inland (o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural (d)</td>
<td>0.512</td>
<td>0.136</td>
<td>0.220</td>
</tr>
<tr>
<td>Coastal (d)</td>
<td>0.310</td>
<td>0.848</td>
<td>0.065</td>
</tr>
<tr>
<td>Inland (d)</td>
<td>0.178</td>
<td>0.015</td>
<td>0.715</td>
</tr>
</tbody>
</table>

(b) Low $\lambda$

<table>
<thead>
<tr>
<th></th>
<th>Rural (o)</th>
<th>Coastal (o)</th>
<th>Inland (o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural (d)</td>
<td>0.628</td>
<td>0.135</td>
<td>0.216</td>
</tr>
<tr>
<td>Coastal (d)</td>
<td>0.147</td>
<td>0.849</td>
<td>0.064</td>
</tr>
<tr>
<td>Inland (d)</td>
<td>0.224</td>
<td>0.016</td>
<td>0.720</td>
</tr>
</tbody>
</table>

(d) Welfare equivalent trade liberalization, low $\lambda$

Table 4: 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.

Under the equalized migration costs, rural migrants are more likely to move to the coastal urban regions, as reported in Panel (a) of Table 4. Unlike the baseline case in Table 1, in the counter-factual simulation 23.9% of the rural population are now moving into the coastal regions, and 13.8% into the inland. This reverses the migration pattern empirically observed as shown in Table 1(b). The re-directed migration flow is approximately 2.9% of the total population. The popularity of the coastal region is expected as it enjoys a higher productivity. As a result of the re-directed migration flow, the population in the coastal urban regions increases by $0.6195/0.5543 - 1 \approx 11.8\%$ while the population in the inland urban area drops by around 6.7%.

The national real wage, which is our measure of national welfare, increases by $0.1922/0.1881 - 1 \approx 2.2\%$ by adopting the $\lambda^*$ policy. The aggregate gain in real wage is substantial, considering the relatively modest change in migration flow of 2.9% of the total population. The real wages also increase in all three regions inside China. The coastal urban region enjoys a higher real wage.

---

9Between the counter-factual and the baseline, $23.9 - 14.9 = 9\%$ of the rural emigrants are re-directed towards the coastal cities. As the initial population in rural China is 0.7080, the re-directed flow is $0.09 \times 0.7080/2.1847 = 2.9\%$ of the total population.
due to the increased population base and market size, which in turn supports a higher number of operating firms (0.0089 v.s. 0.0086) and higher average productivity due to fiercer selection. Even with a reduced population, the inland urban region still enjoys a higher real wage due to the spillover effect from the coastal region. As labor supply concentrates in the high-productivity area, both the intermediate and final consumption goods coming out of the coastal urban become cheaper, and it benefits the inland urban and the rural region through intranational trade.

The changes in urbanization policy also lead to ramifications around the world. At the aggregate level, the other two countries suffer a slightly lower real wage due to the increased import competition from China. After China adopts the $\lambda^*$ policy, the population concentration in the coastal urban region lowers the labor costs there. The Chinese exporters become more competitive in the overseas markets and subsequently drive out the inefficient firms in these countries. As a result of the higher import competition, the number of operating firms in ODC and ROW drops by around 2.2%. Similarly, the improved aggregate productivity in China also means that the Chinese market is harder to export to as well, and correspondingly, the numbers of exporters in the ODC and ROW also decline by around 2%. Besides the negative impacts through import competition, a more productive China can also potentially benefit the other countries. Similar to the impacts on the inland urban area, a larger and more productive coastal region is able to provide the foreign firms with cheaper intermediate inputs, and the foreign consumers cheaper consumption goods. However, these effects on the international stage are much smaller in magnitude due to the presence of trade costs. Consequently, the overall impact on the real wage in the other two countries is negative in this case.

Before moving onto the next exercise, we emphasize that the negative impact in welfare for foreign countries 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. The model, in principle, is rich enough to allow for the urbanization policy in China to lift the welfare of the other countries, as we will see in the next exercise.

\[ 10 \text{Even though the real wage for the rural ROW increases, the overall welfare for the ROW decreases as the rural share of the total population is rather small.} \]

\[ 11 \text{1} - \frac{1.3034}{1.3329} \approx 2.2\% \text{ in the ODC, and } 1 - \frac{40.7147}{41.6206} \approx 2.2\% \text{ in the ROW.} \]
5.1.2 Low $\lambda$ counter-factual

In the second counter-factual exercise, we extend the relatively liberal migration restrictions on rural-inland-urban migration to the rural-coastal-urban migration. In practice, this amounts to setting $\lambda_{31} = \lambda_{21} = 1.79$. The resulting migration pattern is reported in panel (b) of Table 4; the welfare impacts are shown in the last panel of Table 3.

The results are qualitatively similar to the previous exercise with the following notable differences. With the lower migration costs to move out, an additional $0.624 - 0.512 = 11.2\%$ of the rural population migrates out as compared to the baseline and the $\lambda^*$ cases. These additional emigrants prefer to move into the richer coastal urban region, and as a result, the population of the coastal urban region increases by $0.6675/0.5543 - 1 = 20.4\%$ relative to the baseline, and $0.6675/0.6195 - 1 = 7.7\%$ relative to the $\lambda^*$ counter-factual. The further concentration of the workforce in the coastal areas leads to a more substantial welfare gain in all three regions, and a national welfare gain of $0.2011/0.1881 = 6.9\%$. Relative to the $2.2\%$ gain in the $\lambda^*$ counter-factual, the further relaxation of migration restrictions has resulted in a much higher welfare improvement.

In the global context, the negative impacts on the foreign firms are intensified compared to the $\lambda^*$ case: around $1 - 1.0237/1.3329 = 23.2\%$ of the operating firms in ODC, and $1 - 32.1198/41.6206 = 22.8\%$ in ROW are driven out of business. The numbers of exporting firms in the two foreign economies also face a downfall of a similar magnitude. Think conversely: these results echo the idea that the differential reforms that restrain the growth of large, productive, coastal cities may help 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. Moreover, the rural welfare in the ROW improves by $30.8986/29.1714 - 1 = 5.8\%$ while the urban welfare drops by $6.1085/6.0600 - 1 = 0.8\%$. The changes in welfare are mainly driven by the trade in agriculture goods. Under the low $\lambda$ policy, people move out of the rural region in China, and subsequently reduce the agriculture supply; as the migrants earn a higher wage in urban China, the agriculture demand from China surges. The 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 the 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 2b.

### 5.2 Comparing Urbanization Policies with Trade Policies

In the previous subsection we evaluate 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. The urbanization policies and the trade policies target different margins of economic activity: the former concerns the movement of labor while the latter is centered around the mobility of commodities. The comparison between the two puts urbanization policies in a familiar perspective based on the vast literature on trade liberalization. Table 5 summarizes all the results, in which we replicate the baseline results in year 2015 in the first panel for ease of exposition.
The previous subsection has shown that the $\lambda^*$ policy, in which the rural-urban migration friction are equalized while keeping the total emigration of the rural population constant, delivers a 2.2% gain in national welfare. To put this welfare gain into perspective, we solve for a level of trade liberalization that delivers the same 2.2% gain in national welfare. We implement the trade liberalization as uniform reduction in $\tau_{ij}$ 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 5. The goal of such an exercise is to find what the $\lambda^*$ 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 6.9% improvement in national real wage, the same as in the “low $\lambda$” 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 $\lambda^*$ counter-factual, China must reduce the bilateral variable trade costs by 4.4%, and the equivalent of the “low $\lambda$” policy calls for a whopping 12.7% reduction in bilateral iceberg trade costs. 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 the period 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 (c) and (d) of Table 4 to 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, achieving improvements in national welfare by optimizing the spatial

---

12 A 4.4% reduction in bilateral trade costs reduces the iceberg trade costs between China and the ODC from 2.46 to $(2.46 - 1) \times (1 - 0.044) + 1 = 2.392$. Note that the reduction of 4.4% is towards 1, not towards 0, because $\tau_{ij} = 1$ is the limit case of free trade. Similarly, a 4.4% reduction changes the costs between China and the ROW from 2.48 to 2.41. Under the 12.7% reduction, the cost between China and the ODC drops to 2.27, and between China and the ROW, to 2.29.
distribution of population.

Unlike the case of urbanization policy which might hurt foreign countries, bilateral trade liberalization with China benefits both trading partners. Similar to the urbanization policy, reduced trade barriers improve the competitiveness of Chinese firms in the foreign markets. As a result, the numbers of operating and exporting firms decline in all countries. As is common in Melitz models, the intra-industry reallocation induced by trade liberalization improves the aggregate productivity in the differentiated products and eventually leads to gains from trade in both foreign economies.\(^\text{13}\)

### 5.3 Higher Migration Elasticity

<table>
<thead>
<tr>
<th></th>
<th>Rural (o)</th>
<th>Coastal (o)</th>
<th>Inland (o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural (d)</td>
<td>1.00</td>
<td>1.31</td>
<td>1.51</td>
</tr>
<tr>
<td>Coastal (d)</td>
<td>2.06</td>
<td>1.00</td>
<td>2.89</td>
</tr>
<tr>
<td>Inland (d)</td>
<td>1.31</td>
<td>2.41</td>
<td>1.00</td>
</tr>
<tr>
<td>(f_e)</td>
<td></td>
<td></td>
<td>16.70</td>
</tr>
</tbody>
</table>

(a) Joint Calibration of \(\{\lambda_{ij}\}\) and \(f_e, \kappa = 3.3\)

Table 6: Re-Calibrated Parameters Under a Higher Migration Elasticity

Note: this table reports the jointly calibrated parameters when the migration elasticity is set to 3.3. The other parameters as reported in Table 2a and 2b are the same as in the 2015 baseline model.

In this subsection we explore the sensitivity towards the migration of elasticity. In the baseline quantification of the model we have used a migration elasticity of \(\kappa = 1.63\). While our choice of \(\kappa\) 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. The re-calibrated parameters are reported in Table 6 in the Appendix, and the main results of comparing between the baseline and the alternative urbanization policies are reported in Table 7.

A higher migration elasticity implies that the estimated \(\lambda_{ij}\) are smaller in levels and less dispersed as well, as evidenced by comparing the estimates between the robustness and the baseline quantification. Nevertheless, the key pattern is still preserved in the case with higher \(\kappa\): in 2015,\(^\text{13}\)

---

\(^{13}\) Even though the welfare in the rural ROW slightly decreases, the overall welfare in the ROW increases because the rural population is a rather small fraction of the total population.


Table 7: Robustness Check: Higher Migration Elasticity

Note: The 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.

it is significantly harder to move from the rural regions to the coastal urban region ($\lambda_{21} = 2.06$) than to the inland urban region ($\lambda_{31} = 1.31$). This is no surprise, as the relative magnitude of the migration friction is mainly identified via the relative volume of migration flows in the data.

Subsequently, the impacts of the alternative urbanization policies are qualitatively similar but quantitatively larger. Adopting the $\lambda^*$ policy leads to 3.5%, and the “low $\lambda$” policy, a 11.6% increase in national welfare. These numbers are to be compared with the 2.2% and 6.9% welfare gains in the baseline. The welfare gains are higher here because the migration flows are more sensitive to the changes in $\lambda_{ij}$ in a world with a high elasticity.

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 the coastal urban region is higher than that to the inland urban region in 2005, the opposite is true is 2015. Such a reversal in the migration pattern is consistent with the differential reforms on the hukou system that encourage rural people
to move into small and medium sized cities and restrain the growth of large cities, many prominent and productive ones of which are in the coastal provinces.

In our quantitative analysis, we find that equalizing migration costs between rural-coastal-urban and rural-inland-urban migration while keeping the total emigration of the rural population unchanged results in 2.2% welfare gains. The magnitude is substantial considering that this involves a mere 2.9% of reallocation of population from the inland urban to the coastal urban region. We also find that a more laissez-faire urbanization policy that equalizes the rural-coastal-urban migration cost to the rural-inland-urban one results in 6.9% welfare gains. The welfare gains under the two alternative urbanization policies amount to what would result from 4.4% and 12.7% reduction 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 the period when it entered the World Trade Organization (at the end of 2001) and when tariffs were substantially reduced. Namely, China can gain even more from laissez-faire urbanization policies than the trade liberalization that it has accomplished in the past.

Also recall that our welfare results are conservative estimates because of the relatively low migration elasticity adopted in the baseline. As shown in Section 5.3, the welfare gains can 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 or congestion forces in cities besides the idiosyncratic locational preferences being a dispersion force. It would be an interesting extension to incorporate these 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 on the coastal urban region would become all the more important in enhancing aggregate welfare.

References


Appendix

A Supplementary Tables

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

Table A.1: 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](https://www.iso.org/iso-3166-country-codes.html)
B The Model Solution

The equilibrium conditions in the model can be described as a system of nonlinear equations in which \( \{w_j, I_j, P_j, L_j\} \) 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 \( \{w_j, I_j, P_j, L_j\} \) as functions of the current values. The algorithm continues until the current and the implied values of endogenous variables converge under a pre-specified tolerance level, which is 1.0E-6 in this paper. In this appendix, we describe the equations and rules to update each variable above. Before venturing into each variable in detail, we first define some notation and highlight several 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 \( U \), and the rural locations as \( R \), with the understanding that \( U \cup R \) covers all the locations in our model, and \( U \cap R = \emptyset \).

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

\[
\Upsilon_{ij} = (\tau_{ij})^{-\theta} (f_{ij})^{-\frac{\theta-\epsilon+1}{\epsilon-1}}.
\]

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

\[
G_j(a) = \mu_j a^\theta
\]

\[
g_j(a) = \theta \mu_j^\theta a^{\theta-1}.
\]

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 simply the labor income, \( w_j L_j \). The total income in the rural regions adopts the same expression due to the perfectly competitive agriculture market.
Expenditure Out of the total income, $1 - \alpha$ fraction is spent on the differentiated goods by the consumers. Moreover, in the urban region, the firms also demand the differentiated products as inputs. As a result, the total expenditure on the differentiated products, $X_j$, comes from both parts in the urban locations. In the rural regions, $X_j$ solely depends on consumer demand. As a result, we can express $X_j$ as

$$X_j = \begin{cases} (1 - \alpha) w_j L_j + (1 - \beta) X_j = \frac{1 - \alpha}{\beta} w_j L_j, & \text{if } j \in \mathcal{U} \\ (1 - \alpha) w_j L_j, & \text{if } j \in \mathcal{R}. \end{cases} \quad \text{(B.1)}$$

The expenditure on the agriculture products, $X^A_j = \alpha w_j L_j$ in all the locations.

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

$$P_j = \left[ \sum_{i \in \mathcal{U}} \left( \frac{\varepsilon}{\varepsilon - 1} \tau_{ji} \chi_i \right)^{1 - \varepsilon} I_i \int_0^{a_{ji}} a^{1 - \varepsilon} g_i(a) da \right]^{\frac{1}{1 - \varepsilon}}$$

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

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

therefore the rule to update $P_j$, conditional on $X_j$, $I_i$, and $\chi_i$ is

$$P'_j = \frac{\varepsilon}{\varepsilon - 1} \left( \frac{\theta}{\theta - (\varepsilon - 1)} \right)^{-\frac{1}{b}} \left( \frac{X_j}{\varepsilon} \right)^{-\frac{\theta - (\varepsilon - 1)}{b(\varepsilon - 1)}} \left[ \sum_{i \in \mathcal{U}} I_i \left( \frac{\mu_i}{\tau_{ji} \chi_i} \right)^\theta \left( \frac{1}{\chi_i f_{ji}} \right)^{\frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}} \right]^{-\frac{1}{b}}. \quad \text{(B.2)}$$
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

\[ P'_c = \frac{\varepsilon}{\varepsilon - 1} \left( \frac{\theta}{\theta - (\varepsilon - 1)} \right)^{-\frac{1}{\theta}} \left( \frac{X_c}{\varepsilon} \right)^{-\frac{\theta}{\varepsilon - 1}} \left[ \sum_{i \in \mathcal{U}} I_i (\Upsilon_{ci}) (\mu_i)^{\theta} (\chi_i)^{-\frac{\theta - \theta - (\varepsilon - 1)}{\varepsilon - 1}} \right]^{-\frac{1}{\theta}}. \] (B.3)

### B.2 Trade Flow

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

\[ X_{ij} = I_j \int_{0}^{a_{ij}} p_{ij}(a)q_{ij}(a) dG_j(a) \]

\[ = I_j \int_{0}^{a_{ij}} \frac{X_i}{(P_i)^{1-\varepsilon}} \left[ \frac{\varepsilon}{\varepsilon - 1} \tau_{ij} X_j \right]^{1-\varepsilon} \int_{0}^{a_{ij}} a^{1-\varepsilon} dG_j(a) \]

\[ = I_j \frac{X_i}{(P_i)^{1-\varepsilon}} \left[ \frac{\varepsilon}{\varepsilon - 1} \tau_{ij} X_j \right]^{1-\varepsilon} \theta \frac{\theta}{\theta - (\varepsilon - 1)} \left( \mu_j \right)^{\theta} \left( a_{ij} \right)^{\theta - (\varepsilon - 1)} \]

\[ = I_j \frac{X_i}{(P_i)^{1-\varepsilon}} \left[ \frac{\varepsilon}{\varepsilon - 1} \tau_{ij} X_j \right]^{1-\varepsilon} \theta \frac{\theta}{\theta - (\varepsilon - 1)} \left( \mu_j \right)^{\theta} \left[ \frac{\varepsilon - 1}{\varepsilon} P_i \right] \left( \frac{X_i}{\varepsilon \lambda_j} \right)^{\frac{1}{\varepsilon - 1}} \]

\[ = I_j \left( \frac{X_i}{(P_i)^{1-\varepsilon}} \right)^{\frac{\theta}{\varepsilon - 1}} \left( \frac{\varepsilon - 1}{\varepsilon} \right)^{\theta} \left( \tau_{ij} \right)^{-\theta} \left( \chi_j \right)^{-\theta - \frac{\theta - \theta - (\varepsilon - 1)}{\varepsilon - 1}} \left( f_{ij} \right)^{-\frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}} \theta \frac{\theta}{\theta - (\varepsilon - 1)} \left( \mu_j \right)^{\theta} \varepsilon^{-\frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}} \]

\[ = I_j \left( \frac{X_i}{(P_i)^{1-\varepsilon}} \right)^{\frac{\theta}{\varepsilon - 1}} \left( \frac{\varepsilon - 1}{\varepsilon} \right)^{\theta} \Upsilon_{ij} \left( \chi_j \right)^{-\theta - \frac{\theta - \theta - (\varepsilon - 1)}{\varepsilon - 1}} \left( \mu_j \right)^{\theta} \varepsilon^{-\frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}} \].

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

\[ X_{ij} = \frac{I_j (X_i)^{\frac{\theta}{\varepsilon - 1}} \left( \frac{\varepsilon - 1}{\varepsilon} \right)^{\theta} \Upsilon_{ij} \left( \chi_j \right)^{-\theta - \frac{\theta - \theta - (\varepsilon - 1)}{\varepsilon - 1}} \left( \mu_j \right)^{\theta} \varepsilon^{-\frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}} \left[ \sum_{k \in \mathcal{U}} I_k (\Upsilon_{ik}) (\mu_k)^{\theta} \left( \chi_k \right)^{-\theta - \frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}} \right]^{-\frac{1}{\theta}}} {\sum_{k \in \mathcal{U}} I_k (\Upsilon_{ik}) (\mu_k)^{\theta} \left( \chi_k \right)^{-\theta - \frac{\theta - (\varepsilon - 1)}{\varepsilon - 1}} X_i}. \] (B.4)
B.3 Updating $w_j$ in the Urban Regions

The total expenditure in an urban region $j$ should equal the sales to all the urban and the rural regions:

$$w_j L_j + (1 - \beta) X_j = \sum_{i \in \mathbb{R}} X_{ij} + \sum_{i \in \mathbb{U}} X_{ij}.$$  

In the above equation, the LHS is the total expenditure, which includes the expenditures from the workers $w_j L_j$ and from the firms on intermediate inputs, $(1 - \beta) X_j$. 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 $\sum_{i \in \mathbb{R}} X_{ij} = \alpha w_j L_j$. Substitute this into the overall trade balance condition:

$$(1 - \alpha) w_j L_j + (1 - \beta) X_j = \sum_{i \in \mathbb{U}} X_{ij}.$$  

Substitute in the solution of urban-to-urban trade flows from equation (B.4) and the expression of $X_j$ from equation (B.1):

$$\frac{(1 - \alpha)}{\beta} w_j L_j = \sum_{i \in \mathbb{U}} \frac{I_j (\mu_j)^\theta \tau_{ij}^\theta \theta f_{ij} - \theta (\epsilon - 1)}{\epsilon - 1} \left[ w_j^\beta (P_j)^{1 - \beta} \right] \frac{\theta - (\epsilon - 1)}{\epsilon - 1} \frac{(1 - \alpha) w_i L_i}{\beta}$$

$$w_j L_j = \sum_{i=1}^{J} \frac{I_j Y_{ij} (\mu_j)^\theta (\chi_j) - \theta (\epsilon - 1)}{\epsilon - 1} w_i L_i.$$

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. 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:

\[
 w_j L_j = \alpha \sum_{i \in C} w_i L_i.
\]

In this equation, \( w_j L_j \) is the total income of the rural region, set \( C \) is the set of the regions that belongs to country \( c \), and the RHS of the equation is the total expenditure on agricultural goods of all the regions in country \( c \).

2. If country \( c \) imports agricultural products from country \( d \), then the agricultural wage rates between the two countries must satisfy this equation:

\[
 \frac{w_c}{\mu_c} = \frac{\tau_{cd} w_d}{\mu_d}.
\]

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 \( \frac{w_c}{\mu_c} < \frac{\tau_{cd} w_d}{\mu_d} \) as it would imply that country \( c \) should not import from country \( d \). We cannot have \( \frac{w_c}{\mu_c} > \frac{\tau_{cd} w_d}{\mu_d} \) 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 (9) in the urban area comes down to

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

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

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

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

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

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

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

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

Re-arrange:

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

$$
\left( \frac{1}{\varepsilon} \right)^{\frac{-\theta}{\varepsilon - 1}} \left( \frac{\varepsilon}{\varepsilon - 1} \right) \left( \frac{\varepsilon - (\varepsilon - 1)}{\varepsilon - 1} \right) \mu_j^\theta \left( \chi_j \right)^{\frac{\theta}{\varepsilon - 1}} f_e = \sum_{i=1}^{J} \left( X_i \right)^{\frac{\theta}{\varepsilon - 1}} \left( P_i \right)^\theta \chi_{ij}.
$$
The above equation, for all the regions $j = 1, \ldots, J$, can be written in matrix form:

$$
\begin{bmatrix}
\gamma_{11}(X_1)^{\theta} & \gamma_{21}(X_2)^{\theta} & \cdots & \gamma_{J1}(X_J)^{\theta} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{1J}(X_1)^{\theta} & \gamma_{2J}(X_2)^{\theta} & \cdots & \gamma_{JJ}(X_J)^{\theta}
\end{bmatrix}
\begin{bmatrix}
(P_1)^\theta \\
(P_2)^\theta \\
\vdots \\
(P_J)^\theta
\end{bmatrix}
=
\begin{bmatrix}
\left(\frac{1}{\varepsilon}\right)^{-\theta} \left(\frac{e}{\varepsilon-1}\right)^{\theta} \mu_1^{\theta} (\chi_1)^{\theta} f_e \\
\vdots \\
\left(\frac{1}{\varepsilon}\right)^{-\theta} \left(\frac{e}{\varepsilon-1}\right)^{\theta} \mu_J^{\theta} (\chi_J)^{\theta} f_e
\end{bmatrix}.
$$

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

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

Note that from equation (B.2), we have another solution of price, which we denote as $(P_j)^\theta = DD$.

Combining the two solutions, it is straightforward to see $BB = AA * DD$:
After some manipulation and simplification:

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

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

\[
\begin{bmatrix}
\frac{\theta e}{e-1} \chi_1 f_e \\
\vdots \\
\frac{\theta e}{e-1} \chi_J f_e
\end{bmatrix}
= 
\begin{bmatrix}
\gamma_{11} (\mu_1)^\theta (\chi_1)^{-\frac{\theta e - e + 1}{e-1}} & \gamma_{21} (\mu_1)^\theta (\chi_1)^{-\frac{\theta e - e + 1}{e-1}} & \cdots & \gamma_{J1} (\mu_1)^\theta (\chi_1)^{-\frac{\theta e - e + 1}{e-1}} \\
\vdots & \vdots & \ddots & \vdots \\
\gamma_{1J} (\mu_J)^\theta (\chi_J)^{-\frac{\theta e - e + 1}{e-1}} & \gamma_{2J} (\mu_J)^\theta (\chi_J)^{-\frac{\theta e - e + 1}{e-1}} & \cdots & \gamma_{JJ} (\mu_J)^\theta (\chi_J)^{-\frac{\theta e - e + 1}{e-1}}
\end{bmatrix}
\begin{bmatrix}
\frac{\theta e}{e-1} \chi_1 f_e \\
\vdots \\
\frac{\theta e}{e-1} \chi_J f_e
\end{bmatrix}
\]

\[
= 
\begin{bmatrix}
(X_1) \left[ \sum_{i=1}^{J} I_i (\gamma_{1i}) (\mu_i)^\theta (\chi_i)^{-\frac{\theta - \theta (e-1)}{e-1}} \right]^{-1} \\
\vdots \\
(X_J) \left[ \sum_{i=1}^{J} I_i (\gamma_{Ji}) (\mu_i)^\theta (\chi_i)^{-\frac{\theta - \theta (e-1)}{e-1}} \right]^{-1}
\end{bmatrix}
\]

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

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

46
Denote the LHS vector as
\[ \zeta = \Psi^{-1} \begin{bmatrix} \frac{\theta_{e}}{\varepsilon-1} \chi_{1} f_{e} \\ \vdots \\ \frac{\theta_{e}}{\varepsilon-1} \chi_{J} f_{e} \end{bmatrix}. \]

It is straightforward to see, with the understanding that \( \zeta_j \) is the \( j \)-th element of vector \( \zeta \):
\[
\begin{bmatrix}
X_1 \\
\vdots \\
X_J \\
\end{bmatrix} = \begin{bmatrix}
\sum_{i=1}^{J} I_i (\Upsilon_{1i}) (\mu_i)^{\theta} (\chi_i) - \frac{\theta_{e} - \varepsilon}{\varepsilon - 1} \\
\vdots \\
\sum_{i=1}^{J} I_i (\Upsilon_{Ji}) (\mu_i)^{\theta} (\chi_i) - \frac{\theta_{e} - \varepsilon}{\varepsilon - 1} \\
\end{bmatrix}
= \begin{bmatrix}
\Upsilon_{11} (\mu_1)^{\theta} (\chi_1) - \frac{\theta_{e} - \varepsilon}{\varepsilon - 1} & \Upsilon_{12} (\mu_2)^{\theta} (\chi_2) - \frac{\theta_{e} - \varepsilon}{\varepsilon - 1} & \cdots & \Upsilon_{1J} (\mu_J)^{\theta} (\chi_J) - \frac{\theta_{e} - \varepsilon}{\varepsilon - 1} \\
\vdots & \vdots & \cdots & \vdots \\
\Upsilon_{J1} (\mu_1)^{\theta} (\chi_1) - \frac{\theta_{e} - \varepsilon}{\varepsilon - 1} & \Upsilon_{J2} (\mu_2)^{\theta} (\chi_2) - \frac{\theta_{e} - \varepsilon}{\varepsilon - 1} & \cdots & \Upsilon_{JJ} (\mu_J)^{\theta} (\chi_J) - \frac{\theta_{e} - \varepsilon}{\varepsilon - 1} \\
\end{bmatrix}
\begin{bmatrix}
I_1 \\
\vdots \\
I_J \\
\end{bmatrix}
= \Psi' \begin{bmatrix}
I_1 \\
\vdots \\
I_J \\
\end{bmatrix}.
\]

From the last line the solution of the vector \( I_j \) follows:
\[
\begin{bmatrix}
I_1 \\
\vdots \\
I_J \\
\end{bmatrix} = (\Psi')^{-1} \begin{bmatrix}
X_1 \\
\vdots \\
X_J \\
\end{bmatrix}.
\] (B.5)

**B.6 Updating \( L_j \)**

\( L_j \) is directly updated using equation (10), conditional on the solution of \( w_j, P_j, \) and \( I_j \).

**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 $\tau_{ij}$.

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.1 Using the variable names from the dataset, we compute
\( \hat{\tau} \) 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 \( (3.605 - 1) \times (1 - 0.051) + 1 = 3.471 \) in 2006.

C.2 Data Sources, China

**The Input-Output Table of China** We use the 2002 Input-Output Table of China to estimate the agriculture share in consumption (\( \alpha \)) and the labor share in differentiated products (\( \beta \)). 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), and we define industries 02 to 21 as the differentiated industries.

**The 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 2005 survey (Question 11), or was entitled to contract rural land (\( Tu Di CHeng Bao \)) in 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 $j$.

**The 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*)”.

**The Population Census** The Population Census in 2000 and 2010 was used to construct the initial population distribution in the 2005 and the 2015 calibration, respectively. We use the urban population (*Shi Xia Qu Ren Kou*) to measure the relative urban population between the coastal and the inland urban areas. The rural population in China in our paper does not depend on the population census, but on the total population from PWT and the employment in agriculture from WDI.

**C.3 The estimation of $\mu_j$ at the prefecture-level**

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

$$
\log(w_j) = \delta_0 + \delta_1 \log(L_j) + \delta_2 \log(MA_j) + \nu_j,
$$

where $L_j$ is the population of city $j$ and $w_j$ is approximated by the per capita GDP. Both the population and the GDP data come from the City Statistical Yearbooks. The $MA_j$ is a term that 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 the $MA_j$ term can be written as

$$
MA_j = \sum_{i=1}^{J} w_i L_i (\tau_{ij})^{-\theta},
$$
where $\theta$ is the trade elasticity. This term captures the ease of access to markets given a trade cost matrix summarized by $\tau_{ij}$. 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.

We cannot directly use the estimated city-level productivity terms from Ma and Tang (2020b) as their paper uses a different trade elasticity. Instead, we take the $\tau_{ij}$ matrix between the 279 prefecture-level cities from their paper, and recompute the $\text{MA}_j$ terms and re-run the regression with $\theta = 4$, the trade elasticity in this paper. This exercise produces the city-level productivity that we use to infer the relative region-level productivity.

Lastly, note that the above regression excludes the foreign economies. The exclusion is due to two reasons. The first reason is the data limitations. The data on internal trade costs ($\tau_{ij}$) in China is scarce, and the most detailed matrix from earlier work does not have information on the trade costs with the foreign economies. Moreover, the second reason is the inconsistency in the unit of observation. While the data points within China are at the city-level, the foreign economies in this regression would have been countries or even groups of countries. For this reason, the foreign economies in this regression will 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 prefecture-level cities in the reduced-form regression.