Online Appendix for “Urbanization Policy and Economic Development: A Quantitative Analysis of China’s Differential Hukou Reforms”

Wen-Tai Hsu* Lin Ma†

December 2020

A Supplementary Tables

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

Table A.1: Cities in the MUR

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

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

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

Table A.2: Country List

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

B Model Solution

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

Notations

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

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

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

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

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

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

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

Land Rents Cost minimization in the agriculture sector implies that:

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

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

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

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

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

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

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

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

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

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

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

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

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

B.1 Updating P_j

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

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

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

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

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

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

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

B.2 Trade Flow

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

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

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

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

B.3 Updating w_j in the Urban Regions

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

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

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

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

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

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

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

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

B.4 Updating w_j in the Rural Regions

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

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

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

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

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

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

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

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

Simplify:

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

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

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

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

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

B.5 Updating I_j

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

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

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

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

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

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

Re-arrange:

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

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

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

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

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

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

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

After some manipulation and simplification:

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

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

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

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

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

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

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

Denote the LHS vector as

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

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

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

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

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

B.6 Updating N_j

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

B.7 Fixed Entry

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

and selling to i with a productivity a is:

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

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

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

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

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

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

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

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

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

which leads to

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

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

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

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

where

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Substitute in the modified expression of T_c:

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

where:

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

Simplify the solution:

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

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

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

C Data and Quantification

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

C.1 Data Sources, Global

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

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

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

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

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

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

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

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

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

C.2 Data Sources, China

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

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

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

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

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

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

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

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

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

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

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

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

C.3 Estimation of μ_j at the city level

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

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

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

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

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

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

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

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

D Robustness Checks

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

D.1 Higher Migration Elasticity

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

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

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

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

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

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

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

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

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

(d) Fixed Entry

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

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

D.2 Higher Elasticity of Substitution

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

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

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

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

(a) Higher Migration Elasticity

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

(b) Higher Elasticity of Substitution

Table D.2: Robustness Checks: Results I

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

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

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

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

(a) Higher Expenditure Share of Land Consumption

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

(b) Fixed Entry

Table D.3: Robustness Checks: Results II

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

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

D.3 Higher Expenditure Share of Land Consumption

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

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

(a) λ^*, High Migration Elasticity

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

(b) Low λ, High Migration Elasticity

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

(c) λ^*, High Elasticity of Substitution

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

(d) Low λ, High Elasticity of Substitution

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

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

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

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

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

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

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

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

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

(i) λ^*, Fixed Entry

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

(j) Low λ, Fixed Entry

Table D.4: Robustness Checks: Migration Probability Matrices

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

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

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

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

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

D.4 Fixed Entry

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

Appendix B.7.

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

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

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

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

E Additional Results

E.1 Reverting Productivity and Amenity

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

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

(a) Reverting μ_i

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

(b) Reverting ϕ_i

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

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

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

E.2 The Role of the Fixed Exporting Barriers

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

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

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

(a) Welfare

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

(b) Migration Probability, λ^{*}

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

(c) Migration Probability, low λ

Table E.2: Lowering the Fixed Costs of Exporting

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

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

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

References

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

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

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

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

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

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

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