Internal Geographic, Labor Mobility, and the Distributional Impacts of Trade
Online Appendix (Not for Publication)

Jingting Fan *
University of Maryland

Contents

*Department of Economics, University of Maryland. Email address: fan@econ.umd.edu

A Theory Appendix

A.1 Deriving Equation (8)

$$
\begin{aligned}
\pi^e_{o,d} &= \Pr(\frac{v^e_d z_d}{d_{o,d}} \geq \frac{v^d_g z_g}{d_{o,g}}, \forall g \in G) \\
&= \Pr(z_g \leq \frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}} z_d, \forall g \in G,) \\
&= \int_0^\infty \Pr(z_g \leq \frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}} z_d, \forall g \in G | z_d) f(z_d) dz_d \\
&= \int_0^\infty F_d(\frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_1}{d_{o,1}}} z_d, \frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_2}{d_{o,2}}} z_d, ... \frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}} z_d, ...) dz_d,
\end{aligned}
$$

Where $F_d(\frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_1}{d_{o,1}}} z_d, \frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_2}{d_{o,2}}} z_d, ... \frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}} z_d, ...) := \frac{dF}{dz_d}|_{z_g = \frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}} z_d, \forall g \in G}$ is the probability that the draw from region d is z_d *and* this draw dominates all other draws.

Use the functional form of F, it follows that

$$
\begin{aligned}
\pi^e_{o,d} &= \int_0^\infty \exp(-(\sum_{g \in G}(\frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}} z_d)^{-\epsilon_e})^{1-\rho}) * (1-\rho)\epsilon_e(\sum_{g \in G}(\frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}} z_d)^{-\epsilon_e})^{\rho} z_d^{-\epsilon_e - 1} dz_d \\
&= \frac{\int_0^\infty d\exp(-(\sum_{g \in G}(\frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}} z_d)^{-\epsilon_e})^{1-\rho})}{\sum_{g \in G}(\frac{\frac{v^e_d}{d_{o,d}}}{\frac{v^e_g}{d_{o,g}}})^{-\epsilon_e}} \\
&= \frac{(\frac{v^e_d}{d_{o,d}})^{\epsilon_e}}{\sum_{g \in G}(\frac{v^e_g}{d_{o,g}})^{\epsilon_e}}
\end{aligned}
$$

A.2 Deriving Equation (10)

We first derive the distribution of u^e_o, $u^e_o = \max_{d\in G}\{\frac{v^e_d z_d}{d_{o,d}}\}$,

$$
\begin{aligned}
F_{u^e_o}(u) : &= \text{Prob}(u^e_o \le u) \\
&= \text{Prob}(\frac{v^e_d z_d}{d_{o,d}} \le u, \quad \forall d \in G) \\
&= \text{Prob}(z_d \le \frac{u d_{o,d}}{v^e_d}, \quad \forall d \in G) \\
&= F(\frac{u d_{o,1}}{v^e_1}, \frac{u d_{o,2}}{v^e_2}, ..., \frac{u d_{o,d}}{v^e_d}, ...) \\
&= \exp(-[\sum_{d\in G}(\frac{u d_{o,d}}{v^e_d})^{-\epsilon_e}]^{1-\rho}) \\
&= \exp(-[\sum_{d\in G}(\frac{d_{o,d}}{v^e_d})^{-\epsilon_e}]^{1-\rho} u^{-(1-\rho)\epsilon_e}) \\
&= \exp(-\Phi^{e\,1-\rho}_o u^{-(1-\rho)\epsilon_e})
\end{aligned}
$$

It can be shown that, $\forall d \in G$, the cumulative distribution function of u for workers moving from o, to d, is

$$F_{u^e_{o,d}}(u) = F_{u^e_o}(u) = \exp(-\Phi^{e\,1-\rho}_o u^{-(1-\rho)\epsilon_e}),$$

which is a Frechet distribution with position parameter $\Phi^{e\,1-\rho}_o$ and dispersion parameter $(1-\rho)\epsilon_e$.[1]

$$
\begin{aligned}
E(u^e_o|L_{o,d}) &= \int u dF_{u^e_o}(u) \\
&= \int u d(\exp(-[\sum_{d\in G}(\frac{d_{o,d}}{v^e_d})^{-\epsilon_e}]^{1-\rho} u^{-(1-\rho)\epsilon_e})) \\
&= \int u\epsilon_e(1-\rho)\exp(-[\sum_{d\in G}(\frac{d_{o,d}}{v^e_d})^{-\epsilon_e}]^{1-\rho} u^{-(1-\rho)\epsilon_e})[\sum_{d\in G}(\frac{d_{o,d}}{v^e_d})^{-\epsilon_e}]^{1-\rho} u^{-(1-\rho)\epsilon_e} du \\
&= \int u\epsilon_e(1-\rho)\exp(-\Phi^{e\,1-\rho}_o u^{-(1-\rho)\epsilon_e})\Phi^{e\,1-\rho}_o u^{-(1-\rho)\epsilon_e} du \\
&= -\int \epsilon_e(1-\rho)\exp(-y) y du \quad (\text{change of variable} : y = \Phi^{e\,1-\rho}_o u^{-(1-\rho)\epsilon_e}) \\
&= -\int \epsilon_e(1-\rho)\exp(-y) y d(\frac{y}{\Phi^{e\,1-\rho}_o})^{\frac{1}{-(1-\rho)\epsilon_e}} \\
&= \int \exp(-y) y^{-\frac{1}{(1-\rho)\epsilon_e}} \Phi^{e\,\frac{1}{\epsilon_e}}_o dy \\
&= \Phi^{e\,\frac{1}{\epsilon_e}}_o \Gamma(1 - \frac{1}{\epsilon_e(1-\rho)}) \quad (\text{Definition of Gamma function})
\end{aligned}
$$

[1] This is obtained by showing $F_{u^e_{o,d}}(u) := \text{Prob}(u^e_{o,d} \le u | u^e_{o,d} \text{is the highest}) = \frac{\text{Prob}(u^e_{o,d} \le u, u^e_{o,d} \text{is highest})}{\pi^e_{o,d}} = \frac{\int_0^u F_d(z_d) dz_d}{\pi^e_{o,d}} = F_{u^e_o}(u)$.

A.3 Deriving Equation (34)

For workers staying in their hometown, $u^e_{o,o} = \frac{v^e_o z_o}{d_{o,o}} = v^e_o z_o$, hence the distribution of productivity draws for workers choosing to stay in o is:

$$\begin{aligned} F_{z^e_{o,o}}(z) :&= \Pr(z^e_{o,o} < z) \\ &= \Pr(\frac{u^e_{o,o}}{v^e_o} < z) \quad \text{(using } d_{o,o} = 1) \\ &= F_{u^e_{o,o}}(zv^e_o) \\ &= \exp(-[v_o^{-(1-\rho)\epsilon_e} \Phi_o^{e\,1-\rho}] z^{-(1-\rho)\epsilon_e}), \end{aligned}$$

which is also a Frechet distribution. For different regions, the productivity distribution of stayers there have d-ifferent means, but their dispersions will be the same. Therefore, I regress stayers' log wages on regional fixed effects to net out the different average regional productivity draws and interpret the exponents of the residuals as random draws from a Frechet distribution with dispersion parameter $\epsilon_e(1-\rho)$. The coefficient of variations for this distribution is given by Equation (34).

A.4 Proposition 1

Proposition 1 is used in Section C of this appendix, in estimating migration costs.

Proposition 1 *Given migration costs $\{d_{o,d}\}$, there exists a unique set of $\{v_d\}$ (up to normalization), such that the model-predicted number of workers employed in each region equals that in the data, i.e., $L^e_d = \sum_{o\in\mathbf{G}} \pi^e_{o,d} l^e_o$ is satisfied, where L^e_d is the number of workers working in d (data), l^e_o is the number of workers born in o (data), and $\pi^e_{o,d}$ is the model-predicted probability of workers born in o to move to d.*

Proof The proof follows Michaels et al. (2011) and Lemma 1, Lemma 2 in Ahlfeldt et al. (2012), so I only sketch the key steps here.

Consider Equation (9) in the text

$$L^e_d = \sum_{o\in\mathbf{G}} \pi^e_{o,d} l^e_o,$$

Where L^e_d and l^e_o are data, and $\pi^e_{o,d} = \frac{(\frac{v^e_d}{d_{o,d}})^{\epsilon^e}}{\sum_{g\in\mathbf{G}}(\frac{v^e_g}{d_{o,g}})^{\epsilon^e}}$. Given $\{d_{o,d}\}$, l^e_o, and L^e_d, the only unknowns in this equation is $\{v^e_d\}$. Let $\mathbf{v}^e$ be the vector $(v^e_1, v^e_2, ..., v^e_d, ...)$. Define WD$(\mathbf{v}^e)$ (worker deficits) as

$$\text{WD}(\mathbf{v}^e) = L^e_d - \sum_{o\in\mathbf{G}} \pi^e_{o,d} l^e_o.$$

WD is simply the gap between the number of workers working in region d in the data, and the number predicted by the model. WD$(\mathbf{v}^e)$ is a function of $\mathbf{v}^e$. To prove Proposition 1 we show the following:

1. WD$(\mathbf{v}^e)$ is continuous;

2. WD$(\mathbf{v}^e)$ is homogeneous of degree zero;

3. $\sum_{d\in\mathbf{G}} \text{WD}_d(\mathbf{v}^e) = 0, \forall \mathbf{v}^e \in \mathbf{R}^G_+$

4. $\text{WD}(\mathbf{v}^e)$ exhibits gross substitute property.

It is easy to verify that requirement (1) and (2) are satisfied. Requirement (3) can be shown to be satisfied by noting that $\sum_{d\in G} \pi^e_{o,d} = 1$; requirement (4) can be shown to be satisfied by computing the derivatives directly.

Requirements (1)–(2) guarantee the existence of a solution. The proof is a constructive one: by homogeneous of degree zero, we can normalize $\mathbf{v}^e$ to the simplex $\{\mathbf{v}^e \in R_+ : \sum \mathbf{v}^e = 1\}$. Define $\mathbf{WD}^+ = \max\{0, \mathbf{WD}\}$, and $\mathbf{f}(\mathbf{v}) = \frac{\mathbf{v}+\mathbf{WD}^+}{\sum_{d\in\mathbf{G}} v_d + \sum_{d\in textbfG} \mathbf{WD}(v)_d}$, then $\mathbf{f}$ is a continuous function mapping the unit simplex onto itself. The existence of a solution to $\mathbf{v} = \mathbf{f(v)}$ then follows from the Brouwer's fixed point theorem.

Requirement (3)-(4) then guarantee the uniqueness of the solution, see Ahlfeldt et al. (2012) for a more detailed explanation.

The implication of proposition 1 is that, given migration costs, we can solve Equation (9) for the unique set of amenity-adjusted real wages for all locations.

B Background Information and Data Appendix

B.1 Background Information on the Chinese Hukou System

The Hukou system, the history of which dates back to the 1950, is the household registration system in China. It was originally established to control the rural-urban migration in China (back then, residents in cities were subsidized with downward-distorted prices for agricultural products, and there was a strong incentive for people to live in cities). There are two types of Hukous, one for rural residents, the other for urban residents, in each city.[2] Before 1978, people were tied to where their Hukous were and were not allowed to move to any other places without permission from the authority. As a result, there were only minimum rural-urban or urban-urban migrant workers.

Although people are free to travel now, the Hukou system is still important for many aspect of life, as it is tied to health care, social insurance, housing, and education, etc. In many aspects, it acts like within country visa system, distorting the free mobility of labor.[3]

B.2 Data Sources and Sample Construction

The primary individual- and firm-level data I use are the following: the 2005 Mini Population Census, the 2000 Population Census, the 2004 Economic Census, and the 2004 Annual Survey of Industrial Production. In addition to these micro data sources, I also use the 2002 inter-regional and inter-sectoral input-output table, and the data from national accounts and provincial statistical yearbooks.

The 2005 Mini Population Census covers 1% of Chinese population. It records individual demographic and employment information. To my knowledge, this is the only data set that provides individual-level income information for the entire country, so I use it to estimate the average income in each region. I also choose 2005 as the benchmark year, as the calibration procedure requires wage information. The sample I use in this paper is a 1% sub-sample of this data set.

[2] Therefore an urban Hukou in Beijing is different from either an urban Hukou in Shanghai or a rural Hukou in Beijing

[3] See the May 6th, 2010 issue of the magazine The Economist, available at http://www.economist.com/node/16058750, for more information about the Hukou system in China.

The 2000 Population Census covers the entire Chinese population. My sample is its 0.095% sub-sample. Respondents in this sub-sample fill a longer form than others, which asks for information on migration, education, occupation, industry, and housing conditions, but unfortunately, not for information on income.

The 2004 Economic Census covers the universe of registered firms. The sample I have access to is its manufacturing sub-sample, with firm-level revenue and employment information.

The 2004 Annual Survey of Industrial Production covers all state-owned enterprises, as well as private enterprises with annual sales over 5 million RMB yuan. Different from the 2004 Economic Census, this data set contains detailed firm-level financial information, rather than only employment and revenue information.[4]

The rest of this section covers details in sample construction.

B.2.1 Wage

There are two types of workers, two types of local labor markets (rural and urban), and N cities in the economy, so in total there are 4N wages (mean wages for skilled and unskilled workers in all regions in the economy) to estimate. The data I use for this purpose is the 2005 mini census.

I estimate the following specification:

$$log(\text{Wage}_{e,i}) = \beta_0 + \beta_1\text{age} + \beta_2\text{age}^2 + \beta_3\text{sex} + \beta_4 I_{\text{Skilled}} * I_{\text{Agriculture}} + \text{F}_i + \text{S}_i * \text{F}_i I_{\text{Skilled}} + \text{A}_i * \text{F}_i I_{\text{Agriculture}},$$

where F_i is the regional fixed effect, $\text{F}_i * I_{\text{Skilled}}$ is the interaction between regional fixed effect and high-skill dummy, and $\text{F}_i * I_{\text{Agriculture}}$ is the interaction between regional fixed effect and a dummy for agricultural sector. In this specification, I restrict the relative skill premium in the agricultural sector (relative to the skill premium in urban sector of the same city) to be the same across cities (β_4 is not city-specific). This choice is constrained by the power of the regression, as in the sample, in many cities, the rural sector only employee a small number of high-skill workers. The omitted group in the regression is the unskilled worker in the urban sector in Beijing, whose average wage is β_0. Average wages for other groups of workers can be calculated as follows:

Table B.1: Average wage for different groups

Education	Sector	Region	Wage
Unskilled	Urban	i	$\beta_0 + F_i$
Unskilled	Rural	i	$\beta_0 + F_i + A_i$
Skilled	Urban	i	$\beta_0 + F_i + S_i$
Skilled	Rural	i	$\beta_0 + \beta_4 + F_i + S_i + A_i$

The output of the regressions are presented in Table (B.2). The signs and magnitudes of coefficients are reasonable. The R^2 of the regression is 0.58, indicating that the regression has a strong explanatory power. Figure (B.1) presents the distribution of the p-values for the fixed effects in the wage regression. The distribution is heavily concentrated around zero (the spike in the figures corresponds to p-value$<$0.0005), suggesting that the fixed effects are very precisely estimated. Figure (B.2) shows the distribution of average wage for different worker groups across regions. Two patterns emerge: first, there is considerable heterogeneity across regions; second, overall, wages are higher for high-skill workers and urban workers. Figures 2(c) and 2(e) in the text cast the estimates for average

[4] The 2004 Economic Census also covers detailed financial information, but I do not have access to other variables.

Table B.2: Wage Regressions

	(1)
	log_wage
Age	0.0327***
	(22.32)
Age_square	-0.000413***
	(-22.70)
Sex	-0.206***
	(-42.36)
Skilled_agri	-0.296***
	(-16.57)
Observations	62138
R^2	0.576

t statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.01

wages of urban low-skill workers and the average urban skill premia over the map of China. The dispersions in Figure(B.2) show up on the map as the difference both across and within geographic areas.

B.2.2 Migration

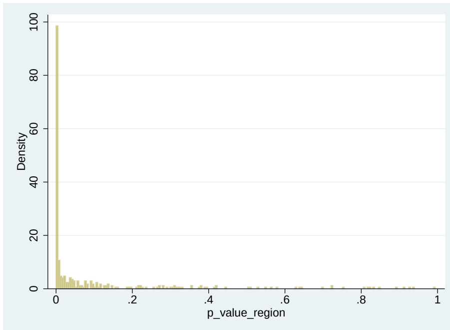
Since I use the 2005 mini census to estimate regional wage and calibrate the model to the 2005 economy, ideally I would like to use this data set to estimate migration costs, too. Since the model neglects dynamic choice of individuals, the migration decision in the model should be best interpreted as a life-time choice. So the model-consistent definition of migration is one that is based on birthplace. However, the 2005 data does not cover birthplace information, so I use the 2000 census to estimate the long-run migration costs.[5] The underlying assumption is that the long-run migration costs do not change much over the period of 2000-2005. It is of course possible that some migration restrictions have been lifted during the period; in that case, the counterfactual experiments in the paper should be interpreted as: what are the welfare implications of international trade for China in 2005, had the migration costs stayed at the 2000 level.

The following are the procedures I use to construct migration flow: first of all, I restrict the sample to those who already finished their schooling, aged between 20 and 60 (60 is the official retirement age for urban male non-physical-labor workers in China), I also drop those who are currently not working, unless the reason for not working is either "on vacation" or "on sick leave". I classify a worker as a migrant, if he or she is not working in her or his birthplace. I identify the source sector (rural or urban) of a worker with the type of Hukou (rural or urban) the worker currently holds, and the destination sector of a worker by the locality the survey respondent.[6] Given the small proportion of workers with college degrees in China in 2000, I classify a worker to be high-skill,

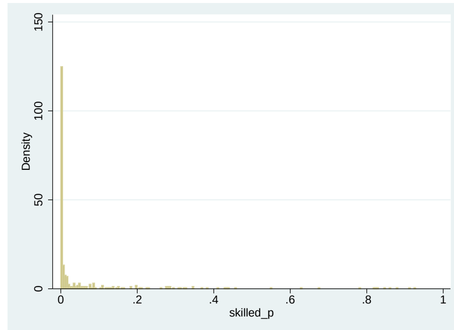
[5] The mini census does report the place of residence in 2000. Therefore one alternative is to combine the migration over the period of 2000-2005 with the long-term migration in 2000, to construct the long-term migration in 2005. This is problematic, as a large fraction of the workers that migrated during 2000–2005 might had been already living outside their birthplace in 2000, i.e., they are repeat migrants. Empirical studies focusing in the U.S. have documented the phenomena of repeat migrant or return migrant (Kennan and Walker, 2011), and the fact that migrants are more likely to respond to economic shocks by migrating, than native workers (Cadena and Kovak, 2013). In light of the evidence, this approach will double count return migrants and repeat migrants, overestimating the long-term migration in 2005.

[6] To the extent that some rural Hukou holders have switched an urban Hukou in 2000, this classification underestimate rural-urban migration. However, until recently, switching a rural Hukou for an urban one was highly restricted.

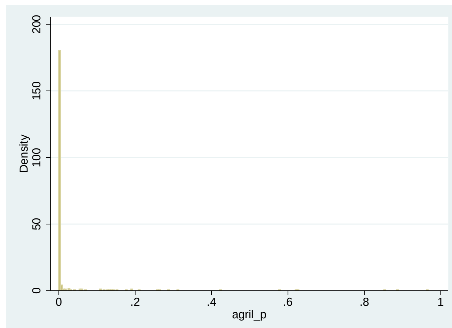
Figure B.1: Distribution of the P-value for Fixed Effects



(a) Regional Fixed Effects

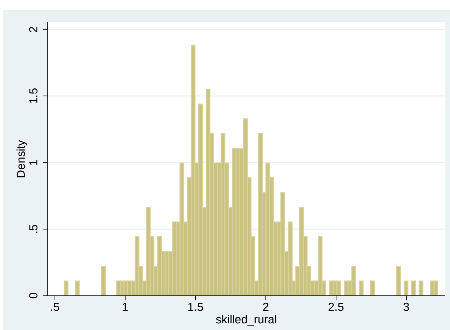


(b) Skilled*Regional Fixed Effects

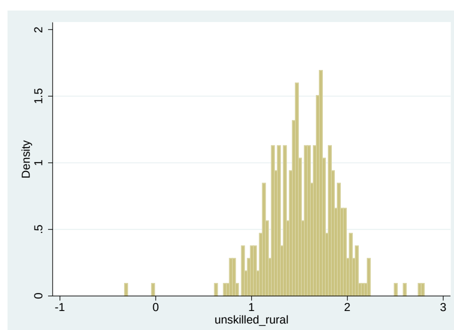


(c) Agriculture*Regional Fixed Effects

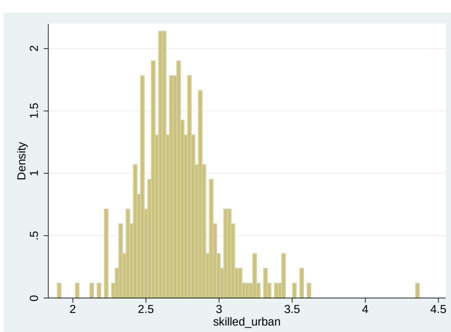
Figure B.2: Average Wages for Different Worker Groups



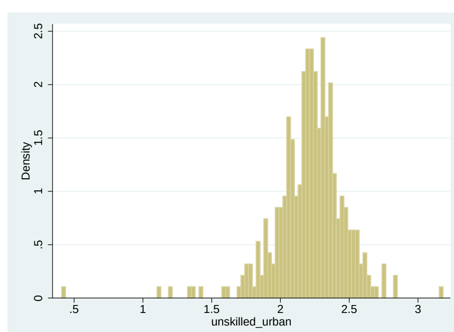
(a) Rural high-skill workers



(b) Rural low-skill workers



(c) Urban high-skill workers



(d) Urban low-skill workers

if he or she has received more than nine years' formal education, equivalent to finishing junior high school.[7] From these procedures, for all workers in the economy, I identify their education level, source province, source sector, destination city, and destination sector. I use this to estimate inter-regional and inter-sectoral migration costs.

B.2.3 Worker Employment and Birthplace Distributions in 2005

Recovering $\{v_d^e\}$ After estimating the parameters governing migration costs, I solve the labor market clearing conditions (Equation 9 in the text) for one more time, to obtain $\{v_d^e\}$ for 2005, the regional fixed effects that are consistent with employment distribution in 2005. For this purpose, I need workers' birthplace and employment distributions in 2005, by workers' level of skills.

I construct the employment distribution from the 2005 mini census. For some cities, due to the small sample size and the small share of skilled workers, there are few skilled workers sampled. For these cities, I supplement the employment distribution aggregated up from the micro data with the published aggregate city-level statistics on employment from the same survey.

I construct workers' birthplace distribution from the 2000 census. I restrict the sample to workers aged 15–55 in 2000. The distribution of this sample will be the distribution for workers aged 20–60 in 2005. To determine the skill level of workers for this sample, if a worker has finished schooling in 2000, I classify his or her skill level based on the education attainment directly; for workers that are above 15, but have not yet finished schooling, I assume they are skilled—by this age, a typical Chinese kid has received 8-9 years of education, so the possibility of (wrongly) classifying a student receiving less than 9 years education as skilled is minimized.

Recovering $\{T_d^s\}$ The employment distribution constructed above gives us the *number of workers* employed in each region. Once we have the estimates for migration costs and regional amenity-adjusted real wages, we can use Equation (12) in the text to convert these into the employment of *effective labor units*. Since there are three industries in urban regions, we still do not know the distribution of employment across industries in each urban region, which is needed for the calibration of productivity at city-industry level.[8] I supplement the regional employment information with the share of employment in industry K over industry M, constructed from the manufacturing sub-sample of the 2004 economic census, and use the service market clearing conditions to obtain the employment information at the city-industry level.[9]

Specifically, let $E_{d,s}^h$ and $E_{d,s}^l$, $s \in \{A, M, K, S\}, d \in \{\mathbf{U}, \mathbf{R}\}$ be the sectoral effective labor unit employment, then regional labor market clearing conditions are:

$$\begin{aligned} &E_{d,A}^h = E_d^h,\ E_{d,A}^l = E_d^l,\ d \in \mathbf{R} \\ &E_{d,M}^h + E_{d,K}^h + E_{d,s}^h = E_d^h,\ E_{d,M}^l + E_{d,K}^l + E_{d,S}^l = E_d^l,\ d \in \mathbf{U} \end{aligned} \tag{B.1}$$

The right sides of these equations are already constructed from the data. Since only agricultural industry is located in rural regions, from the above equation we know labor effective unit employment in the agricultural industry.

[7]The higher education reform started in 1999 in China, which expanded the scale of the higher education sector dramatically. Before the reform, the college admission rate in China was below 5%; in 1999, the college admission increased by 40%. The following years saw additional increase. But until 2005, college graduates constitute only a small proportion of the Chinese labor market.

[8]In the main text, I analyze the intuition behind the quantification strategy in the context of a linear-regression setup, where we need the trade flows between cities for estimation. Such data is not available, so I use a joint quantification strategy, discussed in section C of this appendix, for which I need employment distribution in each city-industry to determine the corresponding productivity.

[9]I do not directly use the 2005 mini census to construct industry-level employment because due to the limited sample size, in some cities, there are no or only a small number of high-skill employment in the capital and equipment industry.

From the optimality conditions of intermediate variety producers, given by Equation (18), the production of intermediate varieties in each place can be calculated, and this should equal to the total demand, D_d^s:

$$
\begin{aligned}
D_d^A &= \frac{E_{d,A}^h W_d^h}{\beta_d^h \gamma_A^L} = \frac{E_{d,A}^l W_d^l}{\beta_d^l \gamma_A^L},\ d \in \mathbf{R} \\
D_d^s &= \frac{E_{d,s}^h W_d^h}{\beta_d^h \gamma_s^L} = \frac{E_{d,s}^l W_d^l}{\beta_d^l \gamma_s^L},\ s \in \{M, K, S\},\ d \in \mathbf{U},
\end{aligned}
\tag{B.2}
$$

With $\{D_d^s : s \in \{A, M, K, S\}\}$ we can compute the city-level demand for industry final output in the service sector, which must equal D_d^S,

$$
D_d^S = C_d^S + C_{d'}^S + +D_{d'}^A \gamma_A^S + \sum_{s \in \{M,K,S\}} D_d^s \gamma_s^A, \quad d \in \mathbf{U}, \tag{B.3}
$$

where C_d^S is the urban service consumption in region d; d' indicates the rural region in the same city as urban region d and $C_{d'}^S$ is the service consumption of this rural region. $C_d^S + C_{d'}^S$ is determined directly by workers' wage and employment distribution. Combine Equations (B.1), (B.2) and (B.3),[10] we have a linear equation system, with 4N unknowns: $E_{d,A}^h$, $E_{d,M}^h$, $E_{d,K}^h$, $E_{d,s}^h$, and 3N equations—(B.3) and the subset of (B.1) for high skilled workers. We combine these three equations with one more data moment—regional employment share in capital and equipment (K) versus other manufacturing industries (M), $\frac{E_{d,K}^h}{E_{d,M}^h}$ to solve for employments of effective labor units in all city-industry.

Once we obtain these employments, we can also use Equation (B.2) to compute the production of intermediate varieties in each industry in all cities.

B.2.4 Factor Shares in Equipped Composite Labor

We need the shares of payments to capital, high-skill workers, and low-skill workers in each region, to calibrate the region-specific equipped composite labor production functions. I compute the ratios between payments to high-skill workers over low-skill workers directly from the estimated wages and the distribution of effective labor units, both of which have been constructed previously. I further need the ratio between the payment to capital, and the payment to labor, in each region.

For the urban regions, I use the 2004 Survey of Industrial Production. I aggregate firm-level data to obtain the city-level ratio between wage bill and expenditures on capital and equipment. The firm-level wage bill is the "total salary payments" entry in the data set; the firm-level expenditures on capital and equipment is the "total capital depreciations" entry in the data set. The total depreciations entry includes, in addition to depreciations to capital and equipment, depreciations to properties and buildings. Therefore I adjust for this by subtracting the share of buildings among aggregate tangible fixed capital stock in China in 2004, calculated from the national statistical yearbook. The mean ratio across cities, constructed this way, is similar to the corresponding ratio from the national input-output table for the urban sector.

For the rural regions, since I am not aware of any data sources that contain information on capital share at the regional level, I assume the capital shares are the same for all rural regions and use the national input-output table to determine it.

[10] We use Equation (B.2) to eliminate $E_{d,s}^l$

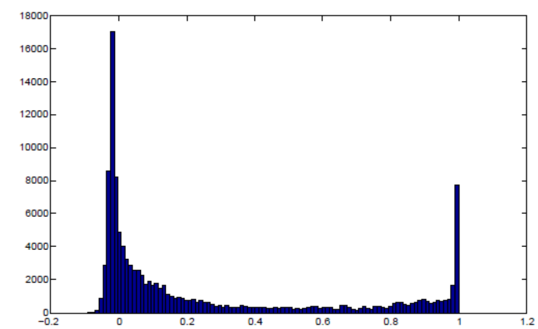
B.2.5 Cultural Distance

To proxy for the cultural distance between cities, I construct a cultural similarity index based on the compositions of ethnic minority groups. I extract the prefecture-level information on the compositions of ethnic minorities from the 1990 census. Migrations was not as pervasive in 1990 as it was in 2000, and therefore the ethnic compositions largely reflect the cultural root of a city. Using the 1990 census data helps us avoid the endogeneity problem that would arise, if we used the 2000 census to construct cultural distance.

There are 56 ethnic groups in China, with Han ethnic being the dominating one. I exclude it, because the share of Han population is so large that including it eliminates most of the variation in the similarity index. For each city, I am left with a 55 by 1 vector, each element of which is the share of one ethnic group in the total local ethnic minority population. I then compute the correlations between the vectors of all city pairs, and use these as the values of my cultural similarity index; the cultural distance is then defined as one minus this similarity index.

Figure (B.3) is the density distribution of the index. The mean, median and standard deviation of the similarity index are 0.2569, 0.0608, and 0.3645, respectively.

Figure B.3: Density Distribution of the Similarity Index



Source: Author's calculation based on the 1990 census

B.2.6 City-level International Trade Surplus

To incorporate international trade imbalances into the calibration, I construct a data set of city-level international trade surplus.

Each city's trade surplus in 2005 is extracted directly from the provincial statistical yearbook. I make two more adjustments. First, Beijing trades a lot with the ROW, but the majority of the trade is done by big companies (especially those SOEs) with headquarters in Beijing. It is plausible that the trade is actually carried in the subsidiaries of these companies, spread out over the country. Fortunately, Beijing statistical yearbook reports "local trade" and "total trade" separately, the later including trade done by SOEs. I assign "local trade" to Beijing, and the remaining component of "total trade" to all Chinese cities, based on their relative size. The implicit assumption is that the operation of those SOEs headquartered in Beijing are distributed across all cities, proportionally to their size.

Second, sometimes the data is not well-behaved. For example, for Shaoshan, a city in Guangdong Province, one of the coastal provinces, the trade surplus is 13 times of its GDP. My conjecture is that there are many trade intermediaries. I make the following adjustments: I aggregate city-level trade surplus to the province level, and then allocate the trade surplus of a province to the cities in the province, according to the GDP of these cities. The underlying assumption is that those trade intermediaries mostly work with other companies in the same province, and trade surplus is proportional to size of economy within a province.

To determine the city-level trade surplus in the scale of the model economy, I first calculate the aggregate trade surplus from the data. I convert the aggregate surplus into the scale of the model and distribute it to all cities, proportionally to each city's contribution to the aggregate trade surplus in the data, constructed above. These are the surplus terms, S_d, in Equation (33).

B.2.7 Input-output Linkages for China and the ROW

In the model, the input-output parameters for China are constructed from the 2002 national input-output table, which records, at the 2-digit industry level, the usages of inputs in the economy. I aggregate the data to four industries—agricultural, capital and equipment, other manufacturing, and service, and four inputs—industry final outputs in the agricultural, other manufacturing, and service industries, as well as equipped composite labor.

The input shares of the ROW are assume to be the same as the median country in Parro (2013). Since the industry classification is finer in this paper, for values not directly available in Parro (2013), I use the corresponding value from China, scaled appropriately. The underlying assumption behind this imputation that, input-output linkages are similar across different countries, are strongly supported by Iones (2013). All results in the paper are robust to changes in the input shares.

Table (B.3) report the shares of inputs in each industry.

Table B.3: Input Shares in China and the ROW

$\gamma_s^{s'}$	Output Industry: China			
Input	A	M	K	S
L	0.57	0.30	0.59	0.48
A	0.19	0.07	0.00	0.03
M	0.15	0.44	0.26	0.21
S	0.09	0.20	0.16	0.28

$\gamma_s^{s'}$	Output Industry: ROW			
Input	A	M	K	S
L	0.58	0.42	0.56	0.63
A	0.19	0.00	0.00	0.00
M	0.16	0.41	0.26	0.11
S	0.07	0.17	0.18	0.26

Notes: This table reports the input shares for different industries in China and the ROW. The source of the values for China is the national input-output table for 2002; the values for the ROW are calculated based on Parro (2013). L stands for the equipped composite labor.

C Estimation and Calibration Appendix

C.1 Calibrating ρ

I obtain an individual panel data from China (China Nutrition and Health Survey), and estimate a Mincer regression with regional fixed effects, along with gender, education, age, and age square as control variables. I then add individual fixed effects to the specification. I compare the R^2 of these two regressions and see how much of the variation unexplained in the first Mincer regression is explained by the individual fixed effects.

As it turns out, about 70% of the unexplained variations can be explained by individual fixed effects. Note that the correlation parameter, ρ, maps one-to-one into the explanatory power of individual fixed effects in the wage regression. For each given value of ρ, I simulate workers' productivity draws from different locations, then estimate a regression specification with *only* individual fixed effects, and calculate the R^2. I chose the correlation parameter so that this R^2 is 70%. This procedure determines a value of 0.4 for ρ.

C.2 Estimating Migration Cost

I use nonlinear least squares to estimate the migration cost, in which $\{\beta\}$ is determined by minimizing the difference between the model-predicted migration flows and their data counterparts. Since the data is at the province-to-city level, I aggregate the predicted city-to-city flows to province-to-city level and take as the objective function the sum of square of the differences between the model's predictions and the data.

Formally, let $p \in \mathbf{P}$ indexes a province in the set of all provinces, $\mathbf{P}$, and $o \in p$ indexes a region o belonging to province p. Recall that l_o^e is the number of workers born in o, and $\pi_{o,d}^e$ is the model-predicted probability for workers to move from o to d, then $l_o^e \pi_{o,d}^e$ is the model-predicted flow from o to d and $\sum_{o \in p} l_o^e \pi_{o,d}^e$ is the aggregate flow from *province* p to *region* d. Let $L_{p,d}^e$ be the flow from p to d in the data, the estimation problem can be formulated in the following way:

$$\min_{\{\beta\}} \sum_{p \in P, d \in G} (\log(\sum_{o \in p} l_o^e \pi_{o,d}^e) - \log(L_{p,d}^e))^2 \tag{C.4}$$

To predict the migration flows using the model, we need to know the regional amenity-adjusted real wages, v_d^e. Because there are more than six hundred regions (rural and urban sectors in 340 cities), it is infeasible to estimate all $\{v_d^e\}$ and $\{\beta\}$ simultaneously. I adopt a nested procedure, similar in spirit to Berry et al. (1995), as follows: in the inner loop, for each given $\{\beta\}$, I solve the migration model for the amenity-adjusted real wages, $\{v_d^e\}$, so that the model-predicted total number of workers in each region is the same as that in data, that is, $\sum_{o \in \mathbf{G}} l_o^e \pi_{o,d}^e = \sum_{p \in P} L_{p,d}^e$, $\forall d \in \mathbf{G}$. Once we have $\{v_d^e\}$, we can compute the model-predicted migration flows, and evaluate the objective function for the given $\{\beta\}$. In the outer loop, I then search over the space of $\{\beta\}$ to minimize the objective function.[11] Proposition 1 in Section A of this appendix ensures the feasibility of this approach by establishing the existence and uniqueness of the solution to the problem in the inner loop.

We use the 2000 migration data, constructed in section A of this appendix, to estimate $\{\beta\}$. After obtaining the estimates, to ensure the recovered $\{v_d^e\}$ are consistent with the 2005 employment distribution, we solve Equation

[11]This nested approach is equivalent to imposing a constraint that the (model-predicted) total numbers of workers migrating to each place equals the total number of workers in that place in the data, and therefore is similar in spirit to what is referred to as "structural gravity estimation" in trade literature. See Fally (2013) for a discussion of the relationship between this and alternative approaches of gravity estimation.

$\sum_{o \in \mathbf{G}} l_o^e \pi_{o,d}^e = \sum_{p \in P} L_{p,d}^e$, $\forall d \in \mathbf{G}$ again, using L_d^e and l_o^e from 2005, to obtain the new $\{v_d^e\}$.

C.3 Jointly Estimating Trade Cost and Productivity

I determine international trade costs, domestic trade costs, and regional productivity jointly.

As discussed in the text, due to the aggregate nature of the data, I use nonlinear least square in estimation, which requires solving the model for the predictions of trade flows. In solving the model, to ensure the size and specialization of the cities in the model are consistent with the data, I compute the production of intermediate varieties in each industries in all cities (details in Section B.2.3 of this appendix), and force the joint estimation algorithm to respect this distribution of intermediate variety production.

Figure (C.4) explains the joint estimation algorithm. I start with an initial guess for international trade costs, and the parameters governing domestic trade costs, $\{\gamma\}$, with which I compute the trade cost between any trade partners, $\{\tau_{o,d}\}$. I then guess a distribution for regional productivity, solve the trade model for prices and trade shares, and check if the demand for intermediate varieties produced by each region equals the supply.[12] If not, I update the guess for the distribution by increasing productivity in regions with excess supply, and decrease productivity in regions with excess demand. The intuition behind this is that, if a region faces excess demand, it means the intermediate varieties produced there is competitive in the international market. To restore the market clearing condition for this region, I make the intermediate varieties produced in that regions more expensive by decreasing the productivity.[13]

Once the distribution of regional productivity that clear all intermediate variety markets are found, I compute the bilateral trade flows, and evaluate the objective function (C.5).

$$\sum_{\text{All P1, P2}} [\log(\frac{X_{P1,P2}}{\text{Domestic Sales}_{P1}}) - \text{the model counterpart}]^2, \tag{C.5}$$

where $X_{P1,P2}$ is the export of goods from province P1 to province P2 in the data. In specifying the objective function, since the domestic trade data is at provincial level, to bring the model and the data together, I aggregate the model-predicted trade flows to provincial level. I normalize the trade flows by aggregate *domestic* sales of the *source* provinces, so that the estimates are not affected by the change in international trade openness between 2002 and 2005.[14]

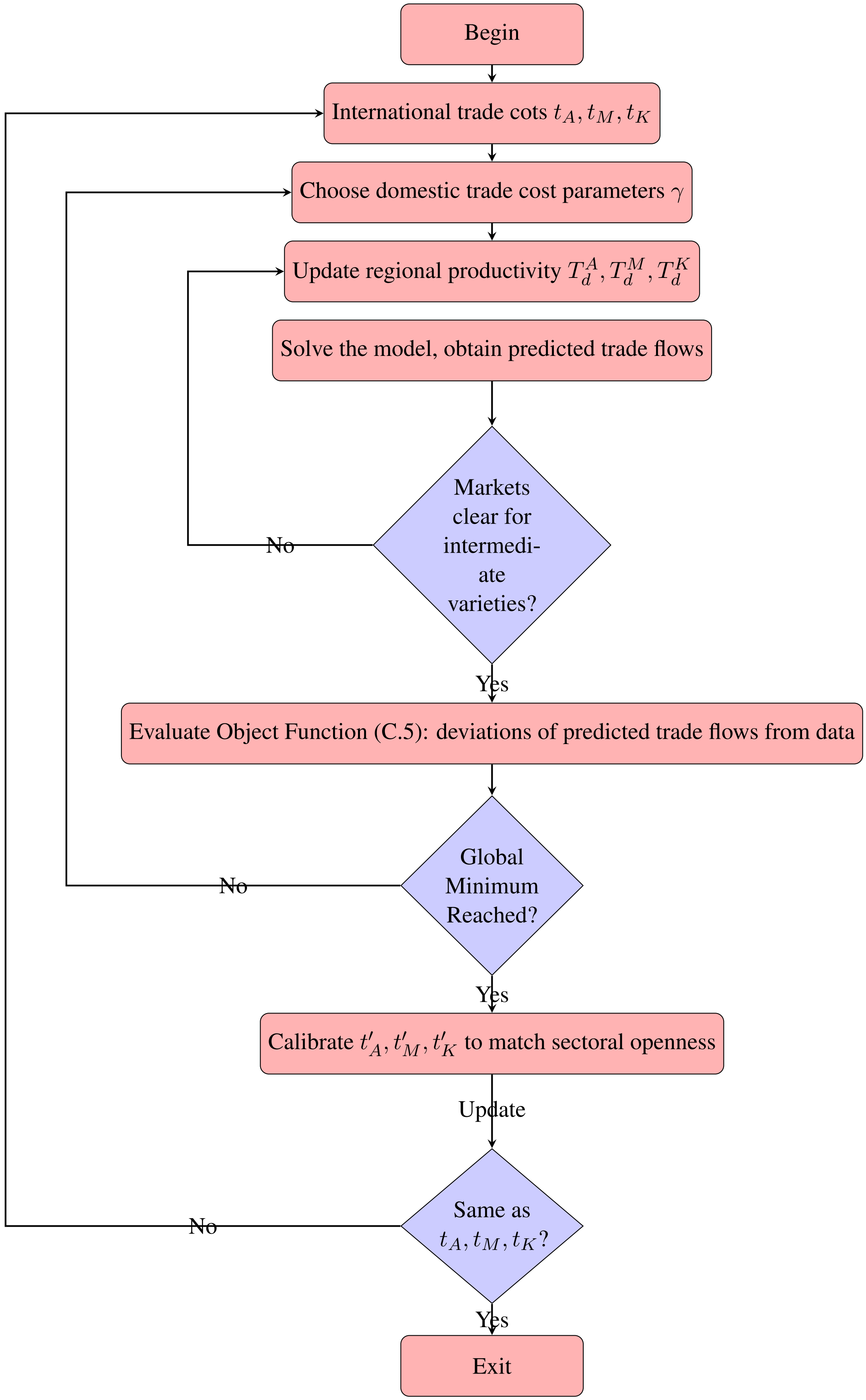
I search over the space of $\{\gamma\}$ until the global minimum is reached, after which I calibrate international trade costs to match the sectoral openness, keeping both domestic trade costs and regional productivity fixed. I repeat the process until convergence.

[12] In the step where we solve the trade model, if we know η_d^h and η_d^s, Equations (17), (25), and (28) in the text can be viewed as a system of equations with prices being the only unknowns. Once we solve these equations for the prices, we can obtain trade shares. Although η_d^h and η_d^s are unknown before the model is parameterized, in section C.4 of this appendix I show that, conditional on information on the shares of different factors in the equipped composite labor, η_d^h and η_d^s are unnecessary in solving the model. Once the model is solved, however, we can use Equation (29) to back out η_d^h and η_d^s, to be used in policy experiments.

[13] The feasibility of this approach requires that, for any given level of trade costs, we can find a set of unique T_d^s that clear all intermediate variety markets in all locations. Redding (2012) proves this is true in a single-sector model. An earlier version of this paper extends the proof to a multi-sector model with input-output linkages within the same broad sector. In the general model here with flexible input-output linkages and capital-skill complementarity, the uniqueness cannot be established. But in implementation, I find the update rule always converge uniformly to one unique object.

[14] The domestic trade data is from 2002, whereas the employment data used to determine production and consumption is from 2005. By normalizing the flows using domestic sales of source provinces, I effectively use only the domestic trade patterns in 2002 for estimation.

Figure C.4: Estimation Algorithm



C.4 Additional Information on the Joint Estimation

In solving the trade model, we need to compute the prices of tradable goods, for the estimated regional wages and given distribution of technology $\{T_d^s\}$. Computing the prices, however, requires η_d^h and η_d^l (see footnote (12)).

To proceed with the estimation algorithm, not knowing η_d^h, η_d^l, I substitute the relative factor shares, $\frac{\text{Capital Share}}{\text{Skilled Share}}$ and $\frac{\text{Equipped Skilled Share}}{\text{Unskilled Share}}$, at the regional level, to the left hand side of Equation(28) in the text, and express η_d^h, η_d^l as

$$\eta_d^h = \frac{(\frac{P_d^K}{W_d^h})^{1-\rho_{kh}}}{\frac{\text{Capital Share}}{\text{Skilled Share}} + (\frac{P_d^K}{W_d^h})^{1-\rho_{kh}}}, \quad \eta_d^l = \frac{(\frac{W_d^{eh}}{W_d^l})^{1-\rho_{lkh}}}{\frac{\text{Equipped Skilled Share}}{\text{Unskilled Share}} + (\frac{W_d^{eh}}{W_d^l})^{1-\rho_{lkh}}} \tag{C.6}$$

I then substitute Equation (C.6) into (28), and solve the model without actually knowing η_d^h or η_d^l. The idea is that, η_d^h and η_d^l must be consistent with the optimal choices of equipped composite labor producers, and therefore when we vary the prices, we also adjust η_d^h and η_d^l so that the optimal factor shares are consistent with data. Once the whole procedure is over and the model is solved, we can then back out η_d^h and η_d^l from (C.6). These are interpreted as the true parameter values, which I keep fixed for all counterfactual experiments.

C.5 Parameters for the Counterfactual Experiments with Different Internal Geographies

In the counterfactual experiments with alternative internal geographies, reported in Section 7.2, I reduce the values of inter-provincial dummies in the trade and migration cost specifications in China to the U.S. level. In this section I describe the sources and values of these parameters.

The value of inter-state trade costs are from Crafts and Klein (2014), which estimates U.S. inter-state trade using the latest data. Under different specifications, their estimates for the inter-state dummy range between 2 to 2.55. To be conservative, I use the upper bound of their estimates, 2.55. This estimate of the inter-state dummy bundles together trade elasticity and trade costs, so I recover the inter-state trade cost by dividing 2.55 by 4, the trade elasticity, arriving at 0.65. Therefore in relevant experiments I reduce the inter-provincial trade costs from the benchmark level of 1.1 to 0.65.

Table C.4: Geographic Parameters in Counterfactual Experiments

		Inter-city (provincial) migration costs	
	Inter-state trade costs	Skilled Workers	Unskilled Workers
Benchmark Level	1.1	1.17 (1.50)	1.20 (1.56)
TC	0.65	1.17 (1.50)	1.20 (1.56)
SMC	1.1	0.99 (0.99)	1.20 (1.56)
UMC	1.1	1.17 (1.50)	0.99 (0.99)

Notes: This table reports the values of the parameters that determine the internal frictions, used in the counterfactual experiments in Table (7). The first column reports the values of the inter-provincial dummy in trade cost specification in different experiments; the second and third columns report the values of inter-city and inter-provincial dummy in migration cost specification, with inter-provincial dummy in parenthesis. Other than the parameters reported in this table, all parameters are kept at their calibrated values.

The value of inter-state migration costs are from Piyapromdee (2014), which estimates migration costs for different demographic groups in the U.S. Since high-skill workers in Piyapromdee (2014) are college graduates, I focus on the low-skill young male group, for which the estimated inter-state migration costs is 99 log points. I apply this value to the inter-provincial migration cost in China, for both unskilled and skilled workers. One further complication is that my specification for migration costs include an inter-city dummy, but the estimated values of the inter-city dummy, for both skilled and unskilled workers, are larger than 0.99. Therefore, in counterfactual experiments, when I reduce the migration costs for certain type of workers, I reduce both the inter-city and the inter-provincial dummy to 0.99.

The parameter values in different cases are summarized in Table (C.4).

C.6 Discussion on the Estimated Inter-Provincial Effect and Additional Robustness

In Section 6.5.2, I report my estimates of the domestic trade costs. It is useful to compare my estimates to those obtained using the U.S. Commodity Flow Survey data. In the literature, the comparable coefficient for state border, after scaled appropriately by the elasticity of trade, is on the range of 0.38 (Wolf, 2000) to 0.65 (Crafts and Klein, 2014, using 2007 data). So my estimate of the state-border effect is about twice as large as the comparable estimates for the U.S., reflecting larger barriers to trade flows at provincial borders in China. One lesson from the U.S. state border literature is that, the estimates might be driven up by the wholesale industry (Hillberry and Hummels, 2003), and might suffer from the aggregation bias—a lot of trade costs are actually due to geographic distance, but might be captured by the state-border dummy when state-level aggregate data is used. When these two factors are taken into account, the estimates shrink (Hillberry and Hummels, 2008).

Therefore, as discussed in Section 6.5.2, one natural concern is whether in China, due to the quality, or the level of aggregation, of the data, the estimates might also misattribute the impacts of geographic distance to the provincial borders; and if that is the case, whether the results from the counterfactual experiments are still valid.

Without detailed micro-level trade flow data available for China, I cannot examine the bias of the estimates. Instead, I use an additional experiment to show that even if there is bias in the estimation, it will not affect main conclusions of the counterfactual experiment. Specifically, I perform a robustness exercise, in which I reduce inter-provincial and inter-regional trade costs to 0.65, the level of the U.S. economy, while at the same time increase the coefficients for the continuous geographic components, so that the overall domestic trade costs and international trade participation are similar to those of the benchmark economy. Effectively, I change the composition of the domestic trade costs, keeping its overall level same as before. I shut down international trade in this economy, and compute the welfare gains from trade, as well as other outcome variables discussed in the text. The results, reported in Table (C.5), are very similar to those of the benchmark experiment, reported in the first column of Table (7).

Table C.5: Counterfactual Experiment with an Alternative Domestic Trade Cost Structure

Panel A: Statistics by Worker Group		
	Mean	std
Urban Skilled	11.52	9.89
Urban Unskilled	5.43	7.41
Rural Skilled	11.10	9.12
Rural Unskilled	5.29	6.74
Panel B: Aggregate Statistics		
National Average	7.47	
Trade Openness	60.60	
Increase in Inequality	6.7	
Contribution-Between(%)	56.88	
Contribution-Within (%)	43.12	

References

Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf, "The Economics of Density: Evidence from the Berlin Wall," *CEP Discussion Paper 1154*, 2012.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, pp. 841–890.

Cadena, Brian C and Brian K Kovak, "Immigrants Equilibrate Local Labor Markets: Evidence from the Great Recession," *NBER Working Paper 19272*, 2013.

Crafts, Nicholas and Alexander Klein, "Geography and Intra-national Home Bias: U.S. Domestic Trade in 1949 and 2007," *Journal of Economic Geography*, 2014.

Fally, Thibault, "Structural Gravity and Fixed Effects," *Working Paper, University of Colorado-Boulder*, 2013.

Hillberry, Russell and David Hummels, "Intranational Home Bias: Some Explanations," *Review of Economics and Statistics*, 2003, *85* (4), 1089–1092.

_ and _ , "Trade Responses to Geographic Frictions: A Decomposition Using Micro-data," *European Economic Review*, 2008, *52* (3), 527–550.

Iones, Charles I, "Misallocation, Economic Growth, and Input-Output Economics," in "Advances in Economics and Econometrics: Tenth World Congress," Vol. 2 Cambridge University Press 2013, p. 419.

Kennan, John and James R Walker, "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 2011, *79* (1), 211–251.

Michaels, Guy, Stephen J Redding, and Ferdinand Rauch, "Technical Note: An Eaton and Kortum (2002) Model of Urbanization and Structural Transformation," *Mimeo*, 2011.

Parro, Fernando, "Capital-skill Complementarity and the Skill Premium in a Quantitative Model of Trade," *American Economic Journal: Macroeconomics*, 2013, *5* (2), 72–117.

Piyapromdee, Suphanit, “The Impact of Immigration on Wages, Internal Migration and Welfare,” *Mimeo*, 2014.

Redding, Stephen J., “Goods Trade, Factor Mobility and Welfare,” *NBER Working Paper 18008*, 2012.

Wolf, Holger C, “Intranational Home Bias in Trade,” *Review of economics and statistics*, 2000, *82* (4), 555–563.