Notes on "Trade, Migration, and Productivity: A Quantitative Analysis of China"

- Paper: Tombe, T., & Zhu, X. (2019). Trade, Migration, and Productivity: A Quantitative Analysis of China. *The American Economic Review*, *109*(5), 1843–1872.
- This paper extends CP2015 and ARSW2015 by adding specific setups (e.g., land institutions) based on China's institutional background.

1. A simple decomposition to calculate the gains from trade and migration

- In this section, the authors develop a simple decomposition to calculate the contribution on the growth of real GDP by different drives.
- Let y_n^j and l_n^j be the real GDP per worker and employment share in region n and sector j, then we have $y=\sum_{n,j} y_n^j l_n^j$. Denote $\hat{y}=y'/y$ to yield:

$$\hat{y}=\sum_{n,j}\omega_n^j\hat{y}_n^j\hat{l}_n^j=1+\sum_{n,j}\omega_n^j g_{y_n^j}+\sum_{n,j}\omega_n^j g_{l_n^j}+\sum_{n,j}\omega_n^j g_{y_n^j}g_{l_n^j}. \tag{1}$$

Proof. For $\hat{y}=y'/y$, we have:

$$\hat{y}=\frac{\sum_{n,j} {y_n^j}' {l_n^j}'}{\sum_{n,j} y_n^j l_n^j}=\frac{\sum_{n,j}\hat{y}_n^j\hat{l}_n^j y_n^j l_n^j}{\sum_{n,j} y_n^j l_n^j}=\sum_{n,j}\omega_n^j\hat{y}_n^j\hat{l}_n^j,$$

where we define $\omega_n^j\triangleq y_n^j l_n^j/(\sum_{n,j} y_n^j l_n^j)$. Then substitute $\hat{y}_n^j=1+g_{y_n^j}$ and $\hat{l}_n^j=1+g_{l_n^j}$ to the above equation to yield:

$$\hat{y}=\sum_{n,j}\omega_n^j(1+g_{y_n^j})(1+g_{l_n^j})=1+\sum_{n,j}\omega_n^j g_{y_n^j}+\sum_{n,j}\omega_n^j g_{l_n^j}+\sum_{n,j}\omega_n^j g_{y_n^j}g_{l_n^j}.$$

■

- Based on the quantitative analysis by Arkolakis et al. (2012), under equilibrium we have $\hat{y}_n^j=\hat{A}_n^j(\hat{\pi}_{nn}^j)^{-1/\theta}$, where A_n^j denotes the labor productivity and π_{nn}^j, π_{nc}^j, and π_{nw}^j denotes the domestic share, other provinces' share, and foreign share of the total expenditure on sector j. Accordingly, we have:

$$g_{y_n^j}\approx g_{A_n^j}-\frac{1}{\theta}\frac{\Delta\pi_{nn}^j}{\pi_{nn}^j}=g_{A_n^j}+\frac{1}{\theta}\frac{\Delta\pi_{nc}^j}{\pi_{nn}^j}+\frac{1}{\theta}\frac{\Delta\pi_{nw}^j}{\pi_{nn}^j}, \tag{2}$$

Proof. The growth of y_n^j can be decomposed $g_{y_n^j} = g_{A_n^j} + g_{(\pi_{nn}^j)^{-1/\theta}}$, where

$$\begin{aligned} g_{(\pi_{nn}^j)^{-1/\theta}} &= \left(\frac{\pi_{nn}^{j}{}'}{\pi_{nn}^j}\right)^{-\frac{1}{\theta}} - 1 \\ &= \left(1 + \frac{\Delta\pi_{nn}^j}{\pi_{nn}^j}\right)^{-\frac{1}{\theta}} - 1 \\ &\approx -\frac{1}{\theta}\frac{\Delta\pi_{nn}^j}{\pi_{nn}^j} \; (\text{when } \frac{\Delta\pi_{nn}^j}{\pi_{nn}^j} \approx 0). \end{aligned}$$

■

- Substitute (1) in (2) and assume $\sum_{n,j} \omega_n^j g_{y_n^j} g_{l_n^j} \approx 0$, equation (1) can be transformed into:

$$g_y = \sum_{n,j} \omega_n^j \frac{1}{\theta} \frac{\Delta\pi_{nc}^j}{\pi_{nn}^j} + \sum_{n,j} \omega_n^j \frac{1}{\theta} \frac{\Delta\pi_{nw}^j}{\pi_{nn}^j} + \sum_{n,j} \omega_n^j g_{l_n^j} + \sum_{n,j} \omega_n^j g_{A_n^j}. \tag{3}$$

- Based on decomposition by equation (3), we can roughly calculate the contribution on the growth by internal trade, external trade, migration, and residual (productivity), with a estimate for θ of 4:

$$g_y = \underbrace{\sum_{n,j} \omega_n^j \frac{1}{\theta} \frac{\Delta\pi_{nc}^j}{\pi_{nn}^j}}_{\text{Internal trade: 4.9\%}} + \underbrace{\sum_{n,j} \omega_n^j \frac{1}{\theta} \frac{\Delta\pi_{nw}^j}{\pi_{nn}^j}}_{\text{External trade: 0.5\%}} + \underbrace{\sum_{n,j} \omega_n^j g_{l_n^j}}_{\text{Migration: 10.8\%}} + \underbrace{\sum_{n,j} \omega_n^j g_{A_n^j}}_{\text{Residual: 40.9\%}}.$$

- One of the important implications from this decomposition is that the internal trade and the migration are larger driving forces than the external trade in China.

2. Quantitative model

2.1. Basic settings

- Perfect competition.
- $N+1$ regions representing China's N provinces plus the world denoted by i and n. In this note, we have slightly modified the notation of the paper and use the script in to represent the trade from i to n .
- Two sectors: agriculture sector and non-agriculture sector denoted by $j, k \in \{ag, na\}$ and use the script kj to represent the migration from k to j.
- Two factor inputs: labor and land.
- Intermediate goods and composite goods keep similar setups to Caliendo and Parro (2015). Intermediate goods can be traded across sections and regions.
- Each worker is registered to a province and assigned either an agricultural or a non-agriculture hukou. Assume workers can move across provinces and sectors within China.

2.2. Worker preferences

- The migration block of the model builds on the work by Ahlfeldt et al. (2015).
- Assume there are $\bar{L}_n^j$ workers with hukou in region n and sector j.
- Define L_n^j as total number of workers in region n and sector j and L_{in}^{kj} as the number of workers with hukou registration in region i and sector k, but works in region n and sector j. Therefore, we have $L_n^j = \sum_{k\in\{ag,na\}} \sum_{i=1}^N L_{in}^{kj}$.

- Worker maximize the C-D type utility function:

$$u_n^j = \varepsilon_n^j \left[\left(C_n^{j,ag}\right)^{\psi^{ag}} \left(C_n^{j,na}\right)^{\psi^{na}} \right]^{\alpha} \left(S_n^{j,h}\right)^{1-\alpha}, \\ \text{s.t. } P_n^{j,ag} C_n^{j,ag} + P_n^{j,na} C_n^{j,na} + r_n^j S_n^{j,h} \le v_{in}^{kj}, \tag{4}$$

  where $C_n^{j,ag}$ and $C_n^{j,na}$ represent the consumption of productions from the agriculture sector and the non-agriculture sector with price $P_n^{j,ag}$ and $P_n^{j,na}$ by region n and sector j, respectively. $S_n^{j,h}$ represents housing structure with price r_n^j. ε_n^j is an idiosyncratic preference variable that is i.i.d. across workers, sectors, and regions.
- The price index for consumers is:

$$P_n^j = \left(\frac{P_n^{j,ag}}{\psi^{ag}} \right)^{\psi^{ag}} \left(\frac{P_n^{j,na}}{\psi^{na}} \right)^{\psi^{na}} \left(\frac{r_n^j}{1-\alpha} \right)^{1-\alpha}.$$

- Define the average income in region n and sector j as $v_n^j = \sum_{k\in\{ag,na\}} \sum_{i=1}^N v_{in}^{kj} L_{in}^{kj} / L_n^j$.
- Since workers' preference are homogeneous, using the property of C-D type utility function, we can derive region n's total demand of goods produced in sector j:

$$D_n^j = \alpha \psi^j \sum_{k\in\{ag,na\}} v_n^k L_n^k, \tag{5}$$

  and n's total demand of the housing:

$$D_n^h = (1-\alpha) \sum_{k\in\{ag,na\}} v_n^k L_n^k.$$

2.3. Production, trade, and good prices

- Composite goods:

$$Y_n^j = \left(\int_0^1 y_n^j(\nu)^{(\sigma-1)/\sigma} d\nu \right)^{\sigma/(\sigma-1)}.$$

- The production function of intermediate goods is C-D type (constant returns to scale) that needs labor input, land input (fixed factor), and composite good input with share of β^j, η^j, and $\sigma^{jk} (k \in \{ag, na\})$. Like Caliendo and Parro (2015), we can derive the price of the input bundle (defined as the term in the bracket):

$$c_n^j(\varphi) = \frac{\Upsilon}{\varphi} \left[\left(w_n^j\right)^{\beta^j} \left(r_n^j\right)^{\eta^j} \prod_{k\in\{ag,na\}} \left(P_n^k\right)^{\sigma^{jk}} \right], \tag{6}$$

  where φ represents the productivity, Υ is a constant that takes the same value across sectors and regions, w_n^j is the wage, r_n^j is the rent of lands, P_n^k is the price of composite goods produced by agriculture and non-agriculture sectors.
- Following the classic setups of EK, assume $\varphi \overset{\text{i.i.d}}{\sim}$ Fréchet distribution with CDF $F_n^j(\varphi) = e^{-T_n^j \varphi^{-\theta}}$, then we can calculate the trade share:

$$\pi_{in}^j = \frac{T_i^j (\tau_{in}^j c_i^j)^{-\theta}}{\sum_{m=1}^{N+1} T_m^j (\tau_{mn}^j c_m^j)^{-\theta}}, \tag{7}$$

where τ_{in}^j is the ice-berg trade cost, and the price index:

$$P_n^j = \gamma \left[\sum_{i=1}^{N+1} T_i^j (\tau_{in}^j c_i^j)^{-\theta} \right]^{-1/\theta}. \tag{8}$$

- Total expenditure on good j by region n is:

$$R_n^j = \sum_{i=1}^{N+1} \pi_{in}^j X_i^j. \tag{9}$$

- Total demand for the good produced in sector j of region n is:

$$X_n^j = D_n^j + \sum_k \sigma^{kj} R_n^k. \tag{10}$$

2.4. Incomes of workers

- Based on China's institutional background, land is not tradable and is owned in common by local residents, which implies that migrant workers have no claim to fixed factor income.
- Since the production function and the preference function are both C-D type, it is easy to derive the total spending on fixed factor is $(1-\alpha)v_n^j L_n^j + \eta^j R_n^j$. Under the profit maximization of producers, labor input satisfies $w_n^j L_n^j = \beta^j R_n^j$, thus the total spending on fixed factor can also be expressed as $(1-\alpha)v_n^j L_n^j + \eta^j (\beta^j)^{-1} w_n^j L_n^j$.
- Given the fixed-factor endowment of $\bar{S}_n^j$, the market clearing condition is:

  $$r_n^j \bar{S}_n^j = (1-\alpha)v_n^j L_n^j + \eta^j (\beta^j)^{-1} w_n^j L_n^j.$$

  Add labor income to both sides of the above equation to yield:

  $$v_n^j L_n^j = (1-\alpha)v_n^j L_n^j + \eta^j (\beta^j)^{-1} w_n^j L_n^j + w_n^j L_n^j.$$

  Thus we can solve the total income under equilibrium:

  $$v_n^j L_n^j = \frac{\eta^j + \beta^j}{\alpha \beta^j} w_n^j L_n^j,$$

  and the total fixed effect income:

  $$r_n^j \bar{S}_n^j = \left[\frac{(1-\alpha)\beta^j + \eta^j}{\alpha \beta^j} \right] w_n^j L_n^j. \tag{11}$$

- Since only workers with local hukou receive fixed-factor income, the income of a local worker in region i and sector j is $w_n^j + r_n^j \bar{S}_n^j / L_{nn}^{jj}$ and the income of a migrant worker is simply w_n^j. Define the effective fixed-factor "rebate rate" to workers:

  $$\delta_{ni}^{jk} = \begin{cases} 1 + \left(\frac{(1-\alpha)\beta^j + \eta^j}{\alpha \beta^j} \right) \frac{L_n^j}{L_{nn}^{jj}} & \text{if } n = i \text{ and } j = k \\ 1 & \text{if } n \neq i \text{ or } j \neq k, \end{cases} \tag{12}$$

  then we can write the incomes of workers registered in region n and sector j as $v_{ni}^{jk} = \delta_{ni}^{jk} w_i^k$.

2.5. Internal migration

- Define the share of workers registered in (n, j) who migrated to (i, k) as m_{ni}^{jk} , where $\sum_k \sum_{i=1}^N m_{ni}^{jk} = 1.$
- Migrant workers face three aspects of migrant costs:
  - Migrants forgo land returns in their home region and rely only labor income.
  - Migrants incur a utility cost that lowers welfare by a factor μ_{ni}^{jk}.
  - Workers differ in their location preferences ε_n^j, which are i.i.d across workers, regions, and sectors.
- Define the real wage as $V_i^k = w_i^k / P_i^k$. Workers from (n, j) choose (i, k) to maximize their welfare $\varepsilon_i^k \delta_{ni}^{jk} V_i^k / \mu_{ni}^{jk}$. Under the law of large number, the proportion of workers who migrant to region (i, k) is:

$$m_{ni}^{jk} = \Pr\left\{\varepsilon_i^k \delta_{ni}^{jk} V_i^k / \mu_{ni}^{jk} \geq \max_{i', k'} \left\{\varepsilon_{i'}^{k'} \delta_{ni'}^{jk'} V_{i'}^{k'} / \mu_{ni'}^{jk'}\right\}\right\}.$$

- Assume $\varepsilon_i^k \overset{i.i.d}{\sim}$ Fréchet distribution with CDF $F_\varepsilon(x) = e^{-x^{-\kappa}}$, where κ governs the degree of dispersion across individuals.
- **Proposition 1.** Given real wage for each region and sector V_i^k, migration costs between all region-sector pairs μ_{ni}^{jk}, land rebate rates δ_{ni}^{jk}, and a Fréchet distribution $F_\varepsilon(x)$ of the heterogeneous preferences, the share of (n, i)-registered workers who migrant to (i, k) is

$$m_{ni}^{jk} = \frac{\left(V_i^k \delta_{ni}^{jk} / \mu_{ni}^{jk}\right)^\kappa}{\sum_{k'} \sum_{i'=1}^N \left(V_{i'}^{k'} \delta_{ni'}^{jk'} / \mu_{ni'}^{jk'}\right)^\kappa} \tag{13}$$

  and total employment at (i, k) is $L_i^k = \sum_j \sum_{n=1}^N m_{ni}^{jk} \bar{L}_n^j$.
  Proof. The proof is similar to that of the "proposition 2" in EK model, so it is omitted here.

2.6. Solving the model

- Exact hat algebra.
- **Proposition 2.** Given changes in migration and real incomes, the change in aggregate welfare is:

$$\hat{W} = \sum_j \sum_{n=1}^N \omega_n^j \hat{V}_n^j \hat{\delta}_{nn}^{jj} \left(\hat{m}_{nn}^{jj}\right)^{-1/\kappa}$$

  where $\omega_n^j \propto \bar{L}_n^j V_n^j \delta_{nn}^{jj} (\hat{m}_{nn}^{jj})^{-1/\kappa}$ is region n and sector j's initial contribution to welfare. Similarly, the change in real GDP is:

$$\hat{Y} = \sum_j \sum_{n=1}^N \phi_n^j \hat{V}_n^j \hat{L}_n^j$$

  where $\phi_n^j \propto V_n^j L_n^j$ is the contribution of region n and sector j to initial real GDP.