Empirical Strategy Brief Notes

James Macek 1001638089

May 25, 2021

1 The model

There are 30 provinces indexed $i \in N$; each can produce goods $s \in \{a, n\}$. Workers in each sector-location are endowed with preferences over goods in these sectors. The utility representation of these preferences is defined implicitly by the relation

$$\sum_{k \in \{a,n\}} \gamma_k^{\frac{1}{\eta}} C_k^{\frac{\eta-1}{\eta}} U^{\epsilon_k \frac{\eta}{\eta-1}} = 1. \tag{1}$$

I normalize the representation so that $\epsilon_n = 1$ and $\gamma_n = 1$ as in Comin et. al (2021). I assume these preferences do not vary across time. Each worker in sector location (i, s) face prices $\mathbb{P}_{i,s}^k$ for goods in sector k. This implies the indirect utility function is defined implicitly by

$$\mathbb{I}_{i,s} = \left[\sum_{k \in \{a,n\}} \gamma_k (\mathbb{P}_{i,s}^k)^{1-\eta} U_{i,s}^{(1-\eta)\epsilon_k} \right]^{\frac{1}{1-\eta}} \tag{2}$$

where $\mathbb{I}_{i,s}$ is the household income of the worker spent on the consumption of goods. This also implies a relation linking expenditure shares on agriculture $\omega_{i,s}$, utility $U_{i,s}$, income and agricultural prices

$$\omega_{i,s} = \gamma_i \left[\frac{\mathbb{P}_{i,s}^a}{\mathbb{I}_{i,s}} \right]^{1-\eta} U_{i,s}^{(1-\eta)\epsilon_a} \tag{3}$$

Each sector good C_k is a CES aggregate of goods sourced by other locations. Each location-sector produces and sells the good at marginal cost $c_{i,k}$. Let θ_k be the trade elasticity corresponding to sector k. I represent the price index of

C_k faced by workers (i, s) as

$$\mathbb{P}_{i,s}^k = T_{i,s}^k \left[\sum_{j \in N} \tilde{\gamma}_i (\tau_{j,i}^k)^{-\theta_k} c_{i,k}^{-\theta_k} \right]^{-\frac{1}{\theta_k}} \tag{4}$$

where $\tilde{\gamma}_i$ are time-constant preference parameters and $\tau_{j,i}^k$ are sector specific trade costs. Since each province represents a large spatial area, I want to capture the empirical regularity that agricultural goods are relatively more expensive to non-agricultural workers if they are more likely to reside in cities– which is crucial for identifying changes in real income, and thus migration. I however, do not have trade information that is dis-aggregated by urban and rural areas.

Note that the representation in (4) implies that the sourcing decisions for all workers within a location i will be identical[1]. This allows us to use trade flow data by sector and province, without having to worry which workers from which occupations are sourcing the goods.

1.1 Land and income

Suppose each worker spends fraction ν of income on goods, and the rest on housing. This means that total income will be $\frac{1-\nu}{\nu} \mathbb{I}_{s,k}$. For simplicity (due to the fact that the utility function does not aggregate well), I assume all workers are paid the same nominal income– i.e. migrants are also entitled to rents from land and capital employed in the sector for which they work in. The aggregate production function in (i, s) is given by

$$y_{i,s} \propto A_{i,s} L_{i,s}^{\sigma_s} H_{i,s}^{\kappa_{H,s}} K_{i,s}^{\kappa_{K,s}} y_{i,s}^{\phi_{a,s}} y_{i,n}^{\phi_{n,s}}$$

$A_{i,s}$ will contain an additional term characterizing external increasing returns; it will endogenously depend on employment of labour per unit of land. A large land share κ_H will mute increasing returns if we assume the supply of land in reach region to be fixed and immobile. Workers are paid the value added share of output in (i, s).

$$A_{i,s} = a_{i,s} \left[\frac{L_{i,s}}{H_{i,s}} \right]^{\Omega_k}$$

[1] Given a level of agricultural spending, each worker will spend the same fraction of that income on goods from each location.

2 Calibrating relative productivity growth and trade costs

Productivity calibration assumes output markets are in equilibrium.

$$Y_{i,s;2005} = \sum_{j\in N,k\in\{a,n\}} \left[\omega_{j,s;2005}\mathbb{I}_{j,k;2005}+\phi_{s,k}Y_{j,k;2005}\right] \frac{\pi^s_{i,j;2000}\hat{\tau^s_{i,j}}^{-\theta_s}\hat{c_{i,s}}^{-\theta_s}}{\sum_{l\in N}\pi^s_{l,j;2000}(\hat{\tau^s_{l,j}})^{-\theta_s}\hat{c_{l,s}}^{-\theta_s}} \tag{5}$$

Proposition: Given sector expenditure shares, GDP and estimates of trade cost growth, there exists a unique (up to scale) vector $\hat{c_{i,s}}^{-\theta_s}$ solving (5). I set $\mathbb{I}_{j,k;2005} = (1-\phi_{a,k}-\phi_{n,k})Y_{j,k}$ so (5) satisfies Walras' Law.

Note that we do not target trade flow data in 2005, and use it only to measure the change in trade costs weighted by the trade elasticity $\hat{\tau^s_{i,j}}^{\theta_s}$. I do, however, use agricultural spending shares in the data to solve the equilibrium uniqueness problem (Our preferences do not satisfy gross substitution) and allows for the recovery of productivity *before* needing to estimate ϵ_a and η.

Trade costs estimation follow a procedure in Tombe and Zhu (2019), which in turn mirrors the trade cost asymmetry procedure done by Waugh (2010). Defer this for later.

3 Preference and absolute productivity growth estimation

The procedure in (2) identifies relative productivity growth. But the theoretical mechanism I have in mind concerns *absolute* productivity growth. The data is informative about absolute productivity growth because welfare is related to agricultural spending via equation (3). Normalize $\hat{c_{i,s}}$ for each s so that they lie on the unit simplex. Let G_s be the scale factor that will identify absolute productivity growth in the data for each sector. Then, one can write equation (2) in hat algebra form:

$$\hat{\mathbb{I}_{i,s}} = \left[\sum_{k\in\{a,n\}} \omega_{i,k;2000}(\hat{\mathbb{P}^k_{i,s}})^{1-\eta}\hat{U_{i,s}}^{(1-\eta)\epsilon_k}\right]^{\frac{1}{1-\eta}} \tag{6}$$

where

$$\hat{\mathbb{P}}^k_{i,s} = \hat{T^k_{i,s}} G_s \left[\sum_{j \in N} \pi_{j,i;2000} (\hat{\tau^k_{j,i}})^{-\theta_k} \hat{c_{i,k}}^{-\theta_k} \right]^{-\frac{1}{\theta_k}} \tag{7}$$

I start by using an exact identification procedure to pin down the growth of within region relative prices. Equation (3) implies the following relationship:

$$\hat{T_i}^{1-\eta} = \frac{\hat{U_{i,n}}}{\hat{U_{i,a}}}^{(1-\epsilon_a)(1-\eta)} \left[\frac{\hat{\omega_{i,n}}}{1-\omega_{i,n}}\right]\left[\frac{\hat{\omega_{i,a}}}{1-\omega_{i,a}}\right]^{-1} \tag{8}$$

So T_i, U_i are all estimated based on solving equations (6) to (8) for a given value of G_s. G_s are each chosen to match the aggregate fall in relative agricultural spending for both agriculture and nonagriculture workers.

$$placeholder \tag{9}$$

Then, I choose ϵ_a and η to minimize the squared distance between the predicted value of relative spending in each sector location and the data. In particular,

$$\min_{\epsilon_a, \eta} \sum_{i \in N, s \in \{a,n\}} \left[\left[\frac{\hat{\mathbb{P}}^a_{i,s}}{\hat{\mathbb{P}}^n_{i,s}}\right]^{1-\eta} \hat{U_{i,s}}^{(1-\eta)(\epsilon_a - 1)} - \left[\frac{\hat{\omega_{i,s}}}{1-\omega_{i,s}}\right] \right]^2 \tag{10}$$

subject to (6), (7), (8) and (9).

4 Measuring capital and land allocation