Misallocation under Heterogeneous Markups and Non-Constant Returns to Scale

Xiaoyue Zhang (Tilburg)
Junjie Xia (Peking)

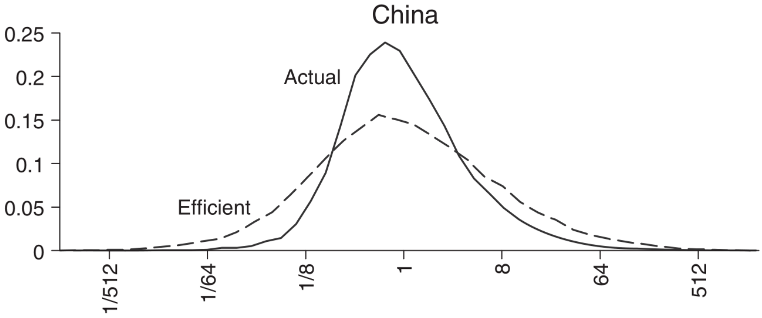
Giuseppe Forte, UCL · June 18, 2022

Hsieh Klenow (2009) in Three Sentences and One Figure

Much more MPL and MPK dispersion in China and India than US.

Model of the economy with CRS firms and a unique elasticity of demand.

Optimally assigning resources (= equalizing TFPR): Chinese TFP $\uparrow$ 86%.



Distributions of Value Added from Hsieh Klenow (2009, Figure III)

Zhang Xia (2022) in Four Sentences and One Table

Introduce flexibility in Hsieh Klenow's model.

Returns to scale of firm production function are not constrained.
Elasticity of demand varies across industries and (less so) within industry.

Optimally assigning resources $\neq$ equalizing TFPR across firms/industries.

Data	α	σ	TFP gains (%)
HK	calibrated using US firms (HK)	3	86.6
HK	calibrated using US firms (HK)	8.5	362.3
HK	calibrated using US firms (HK)	heterogeneous (one type)	298.6
HK	Our estimators	3	51.5
HK	Our estimators	8.5	63.8
HK	Our estimators	heterogeneous (one type)	59.2
Our	Our estimators	8.5	46.3
Our	Our estimators	heterogeneous (two types)	43.9

TFP Gains Across Specifications from Zhang Xia (2022, Table 9)

Models

Hsieh Klenow (2009)

$$Y_{is} = A_{is} K_{is}^{\alpha^K} L_{is}^{1-\alpha^K}$$

$$Y_s = \left(\sum_{i \in s} Y_{is}^{\frac{\varepsilon - 1}{\varepsilon}} \right)^{\frac{\varepsilon}{\varepsilon - 1}}$$

$$\mathscr{Y} = \prod_s Y_s^{\beta_s}, \quad \sum_s \beta_s = 1$$

Zhang Xia (2022)

$$Y_{is} = A_{is} K_{is}^{\alpha_s^K} L_{is}^{\alpha_s^L}$$

$$Y_s = Y_{\bar{s}}^{\gamma_s} Y_{\underline{s}}^{1-\gamma_s}$$

$$Y_{\bar{s}} = \left(\sum_{i \in \bar{s}} Y_{i\bar{s}}^{\frac{\varepsilon_{\bar{s}} - 1}{\varepsilon_{\bar{s}}}} \right)^{\frac{\varepsilon_{\bar{s}}}{\varepsilon_{\bar{s}} - 1}}$$

$$Y_{\underline{s}} = \left(\sum_{i \in \underline{s}} Y_{i\underline{s}}^{\frac{\varepsilon_{\underline{s}} - 1}{\varepsilon_{\underline{s}}}} \right)^{\frac{\varepsilon_{\underline{s}}}{\varepsilon_{\underline{s}} - 1}}$$

$$\mathscr{Y} = \prod_s Y_s^{\beta_s}, \quad \sum_s \beta_s = 1$$

- Firms are price-takers in input markets but price-setters in the output market.
- Single good firms ($\simeq$ sale prices are appropriate weight for ‘aggregate output’).

Why bother with $\alpha^K + \alpha^L \neq 1$ and ε_s?

$$y_{it} = a_{it} + \alpha_s^K k_{it} + \alpha_s^L l_{it} = a_i + \alpha_s^K k_{it} + \alpha_s^L l_{it} + u_{it}^y$$

From CES demand, $y_{it} - y_{st} = -\varepsilon_s (p_{it} - p_{st})$.
From the definition of deflated revenue, $r_{it} = y_{it} + p_{it} - p_{st}$.

The three together imply

$$\Delta r_{it} = \frac{\varepsilon_s - 1}{\varepsilon_s} \alpha_s^K \Delta k_{it} + \frac{\varepsilon_s - 1}{\varepsilon_s} \alpha_s^L \Delta l_{it} + \frac{1}{\varepsilon} \Delta y_{st} + \frac{\varepsilon_s - 1}{\varepsilon_s} \Delta u_{it}^y$$

Finding: returns to scale $<< 1$; significant ε_s variation.
Rest of the paper: convincingly show that this matters.

- Main text only contains $\{\alpha^K, \alpha^L\}$; could include $\{\alpha_s^K, \alpha_s^L\}$ where feasible.
- Is it just fine to treat gross value added like revenue?
- Data cover the period of China WTO accession: should ε_s change over time?
- Do you have industry production data? Cannot replace y_{st} with r_{st}.

Problems: Step 1

> *"the assumption of variable input is very restrictive and it is almost impossible to find a truly variable input in data [. . .] Since labor and capital are far from being variable, applying the production approach in our case is problematic. [. . .]"*

But capital is assumed flexible in anything that follows from profit FOCs.

Step 1 uses $\frac{\text{Cost}}{\text{Revenue}}$ from $\min_{K_i,L_i} \mathscr{C}$ under capital flexibility.

Similarly step 3 of the estimator is invalid if capital is predetermined.

Authors seem to argue "accounting" approach suffers from fewer problems than "production" approach.
But firm problems are dual.

Problems: Step 1

i chooses $\{K_i, L_i, P_i\}$ knowing $\{A_i, \tau_i^L, \tau_i^K\}$; then δ_i is realised.

In $\min_{K_i, L_i} \mathscr{C}$ FOCs, using $\lambda = \mathbb{E}MC$:

$$s_{ig}^K \equiv \frac{RK_{ig}}{P_{ig}Y_{ig}} = \frac{\mathbb{E}MC}{P_{ig}} \frac{1}{\mathbb{E}e^{\delta_{ig}}} \frac{\alpha_s^K}{1+\tau_{ig}^K} = \frac{\varepsilon_g - 1}{\varepsilon_g} \frac{e^{\delta_{ig}}}{\mathbb{E}e^{\delta_{ig}}} \frac{\alpha_s^K}{1+\tau_{ig}^K} e^{-\delta_{ig}}$$

$$\mathbb{E}\ln s_{ig}^K = \ln \alpha_s^K - \mathbb{E}\ln(1+\mu_{ig}) - \mathbb{E}\ln(1+\tau_{ig}^K) - \mathbb{E}\delta_{ig}$$

Authors assume:

- $\delta_{ig} \sim N(0, \sigma_g^2)$, so $\mathbb{E}\ln(1+\mu_{ig}) = \ln \frac{\varepsilon_g}{\varepsilon_g - 1} + \frac{\sigma_g^2}{2}$ and $\mathbb{E}\delta_{ig} = 0$.
- $\mathbb{E}\ln(1+\tau_{ig}^K) = \mathbb{E}\ln(1+\tau_{ig}^L) = 0$, effectively.

Hence $\mathbb{E}\ln s_{ig}^K - \mathbb{E}\ln s_{ig}^L = \ln \alpha_s^K - \ln \alpha_s^L$. Observed, $\perp g$.
Then, $\ln s_{ig}^K - \ln s_{ig}^L = \mathbb{E}\ln s_{is}^K - \mathbb{E}\ln s_{is}^L + \ln(1+\tau_{ig}^K) - \ln(1+\tau_{ig}^L)$.

Step 1 cannot identify $\{\alpha_s^K, \alpha_s^L\}$ without cross-s or RtS assumptions.
But if that is true, the object we plug into Step 2 is not $\ln(1+\mu_{ig})$.

Problems, and Some Progress

Though the authors acknowledge underidentification (appendix H), it is not quite clear what we are estimating if the parameters could not be estimated with infinite data.

We can attribute sources of underidentification to departures from HK:

- $\alpha_s^K + \alpha_s^L \neq 1$ does not hinder identification of τ_i^K if we use $\{\tau_i^Y, \tau_i^K\}$ as in HK rather than $\{\tau_i^L, \tau_i^K\}$;
- using $\{\tau_i^L, \tau_i^K\}$ rather than $\{\tau_i^Y, \tau_i^K\}$ as in HK does not hinder identification of $\alpha_s^K = 1 - \alpha_s^L$.

This means one can ignore interaction among restrictions:

- simplifying the model in one direction will not complicate the other;
- fixing the model in one direction will not fix the other.

Concluding

The authors make a convincing point that the estimates by Hsieh Klenow hinge on assumptions easily disproved by the data.
They extend the original model well.

However, as it stands the model is underidentified, for $\perp$ reasons.
Many (partial) fixes for these problems:

- giving up $g \in \{\underline{s}, \bar{s}\}$ (Klette Griliches regression);
- giving up $\alpha_s^K + \alpha_s^L \neq 1$;
- giving up $\{\tau_i^L, \tau_i^K\}$ for $\{\tau_i^Y, \tau_i^K\}$;
- and many more I can't see.