Productivity Cost of Input Distortions in China *

Xiaoyue Zhang† Junjie Xia ‡

April 18, 2023

The Latest Version

Abstract

Predicted total factor productivity (TFP) gains in China from removing the input distortions under Hsieh and Klenow (2009)'s framework are, in theory, sensitive to the assumption of constant returns to scale and demand elasticity being 3 for all the firms, and that Chinese firms have the same technology as American firms. However, there is little empirical evidence on how the predicted TFP gains would change if these assumptions are relaxed. Using the framework developed by Zhang and Xia (2023), we find that Chinese firms have constant returns to scale on average, that the average demand elasticity is around 8 with large dispersion, and that American firms are more capital intensive than Chinese firms. The large variation in demand elasticities has a minor impact on predicted TFP gains, but using the estimated average demand elasticity implies more than tripled predicted TFP gains. The technological differences between Chinese and American firms do not affect the predicted TFP gains through the estimated input distortions but through aggregation across firms and through the inferred productivity. Our estimates show a 45% TFP gain from removing the input distortions in China in 2005.

Keywords: Input distortions, heterogeneous demand elasticities, TFP gains

*We are grateful to Jaap H. Abbring, Christoph Walsh, and Jeffrey R. Campbell for their patience, support, and guidance throughout this project. This paper has benefited from insightful discussions with and comments from Daniel Xu, Has van Vlokhoven, Malik Çürük, Guiying Laura Wu, Chang-Tai Hsieh, Giuseppe Forte, Jo Van Biesebroeck, Jan De Loecker, Frank Verboven as well as participants at the Tilburg Structural Econometrics Group, Tilburg University GSS seminars, the IO seminar series at KU Leuven, the ENTER Jamboree, China International Conference in Macroeconomics, the IAAE annual conference, the AMES China meeting, China Center for Economic Research's Summer Institute, the NSE International Annual Workshop, and EEA-ESEM. This project is supported by CentER and IAAE Student Travel Grant. Xiaoyue is also thankful for the never-ending support and confidence from Edi Karni.

†Tilburg University, corresponding author, email: x.zhang_11@uvt.nl

‡Central University of Finance and Economics and Peking University, email: junjiexia@nsd.pku.edu.cn

1 Introduction

Standard competitive-market theory predicts that equalizing the marginal revenues of production factors across firms brings efficiency gains (Melitz (2003), Restuccia and Rogerson (2008)), and an inefficient allocation of production inputs creates dispersion in the marginal revenues. Removing the variation in marginal revenues or equating the variation in developing countries with the one in developed countries provides the predicted TFP gains from removing input distortions. This method was introduced by Hsieh and Klenow (2009) (hereafter HK) and has been widely used to evaluate the impact of input distortions. Intuitively, input distortions are anything that prevents firms from using production inputs at their market prices or from freely adjusting their usage. However, the predicted TFP gains in HK rely on the assumptions of constant returns to scale and all the firms' demand elasticities being 3, and the assumption that a developing country and a developed country share the same technology. Haltiwanger et al. (2018) shows how these assumptions can bias the predicted TFP gains in theory but whether predicted TFP gains are sensitive to these assumptions in practice remains unclear.

These assumptions can bias predicted TFP gains in empirical studies if the assumptions are bad descriptions of empirical data and if the predicted TFP gains are sensitive to the assumptions. More specifically, constant returns to scale is a reasonable assumption if firms have indeed constant returns to scale. The assumption of homogeneous demand elasticity may be too restrictive when firms actually face heterogeneous demand elasticities. However, whether this assumption on demand elasticities causes a large bias in the predicted TFP gains depends on how sensitive the predicted TFP gains are to dispersion in demand elasticities. Broadly speaking, if an assumption, on the one hand, allows a model to be tractable, and, on the other hand, does not cause a large bias in empirical results, it can still be a reasonable assumption. In this paper, we first use estimated demand and production parameters to examine how well the assumptions in HK can describe the Chinese firm-level survey data. We then relax the assumptions one by one to show whether the predicted TFP gains are sensitive to the assumptions. We use the framework developed in and the same data as in Zhang and Xia (2023) (hereafter ZX), where returns to scale, production parameters, and demand elasticities are estimated using Chinese firm-level survey data in 2005, and find that Chinese firms in 2005 have on average constant returns to scale, that the average demand elasticity is around 8 and there is large dispersion in demand elasticities, and that American firms are more capital intensive than Chinese firms. The large variation in demand elasticities has a minor impact on predicted TFP gains, but using the average of estimated demand elasticity triples predicted TFP gains. The technological differences between Chinese and American firms do not affect the predicted TFP gains through the estimated input distortions but through aggregation across firms and the inferred firm productivity. Using ZX's estimates gives a 45% TFP gain from removing the input distortions in China in 2005.

In ZX, the production parameters are estimated using firms' observed factor shares under the assumption that the modes of capital and labor distortions in an industry are both 0, that firms' production functions are Cobb-Douglas, and that firms from the same industry have the same production functions apart from the Hicks-Neutral firm-specific productivity. The last two assumptions on production functions imply that the production elasticities are the same inside an industry. The means that the estimated returns to scale are allowed to differ from 1. Demand elasticities are estimated using observed firm-level revenue and cost data, and demand elasticities are allowed to differ within an industry by allowing each industry to consist of a high-demand-elasticity nest and a low-demand-elasticity nest. The demand structure has nested constant elasticities of substitution (nest CES). The elasticities of substitution between nests are assumed to be 1, whereas the within-nest ones are larger than one and are estimated using firm-level data. We follow ZX by allowing more variation in demand elasticities than production elasticities, i.e. production elasticities are allowed to vary inside an industry while production elasticities are constant, because Autor et al. (2020) find empirical evidence that production inputs are reallocated within industries among firms with different demand elasticities. In theory, allowing these heterogeneous demand elasticities can affect the estimation of the other parameters, including input distortions and production elasticities, and therefore affect the predicted TFP gains. Input distortions are firm specific and are the gap between firms' marginal revenues of an input and the market price of the input. We assume that the input markets are perfectly competitive in the absence of input distortions, in the sense that firms are price takers and face the same market prices.

Using the estimates of ZX, we find that removing the input distortions in China would cause a 45% TFP gain compared to the 87% in HK, among which reallocation within industry leads to a 32% TFP gain while reallocation across industries causes a 9% gain.

We then do experiments where we relax the assumptions imposed in HK one by one. Replacing the value 3 assumed for demand elasticities in HK by the average of ZX's estimates, the predicted TFP gains would increase by more than three times. The change in predicted TFP gains is much smaller when replacing the estimated average by estimated demand elasticities. Depending on what production parameters we use and whether we allow demand elasticities to differ within industries, the change is between $\frac{1}{3}$ and $\frac{1}{10}$ of the predicted TFP gains using the estimated average. This implies that predicted TFP gains in China in 2005 are much more sensitive to the average value of demand elasticities but less sensitive to the dispersion in demand elasticities.

This paper complements the critique of Haltiwanger et al. (2018) by reviewing how to interpret the predicted TFP gains estimated in HK when some of the assumptions imposed in HK are violated. Haltiwanger et al. (2018) criticizes that failing to account for heterogeneous demand elasticities can contaminate the inferred input distortions by variation in demand elasticities. We find that if the errors in estimated input distortions are the same for firms inside an nest, the predicted TFP gains are independent from the errors because only the dispersion of input distortions inside a nest affects the predicted TFP gains not the level of input distortions. Therefore, allowing each nest to have its own demand elasticities does not affect the predicted TFP gains through the value of input distortions. However, the value of the demand elasticities affects how the impact of input distortions is aggregated and consequently affects the predicted TFP gains through the aggregation. Similarly, if the production elasticities are unknown and the estimation of production elasticities requires some knowledge about demand elasticities, failing to account for heterogeneous demand elasticities can affect the estimated production elasticities and therefore the predicted TFP gains through the aggregation but not through the values of the input distortions.

We adopt estimates of production elasticities from ZX instead of estimates or methods of other related studies because allowing firm-specific input distortions implies that firms with higher productivity may face higher distortions, as it is most likely the case in China where domestic private firms are understood to be more productive but underuse inputs due to, for example, financial constraints (Song et al. (2011)). Therefore, we cannot use the method of estimating production functions developed by Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2015), where the monotonicity between unobserved firm productivity and some production inputs is required. We prefer ZX to Ruzic and Ho (2021) because we want to allow for heterogeneous demand elasticities within industries and the possibilities that most or all the firms inside an industry experience positive input distortions, or similarly negative input distortions. When an industry, such as the industry of electrical vehicles, is deemed as strategic by the Chinese government, most or even all the firms in the industry may receive favorable financial credits and therefore the average capital distortions in this industry will be negative. Ruzic and Ho (2021) requires demand elasticities to be constant inside industries and assumes input distortions to be mean zero.

Our paper also contributes to the discussion whether constant returns to scale is consistent with firms' empirical production decisions. While many studies assume constant returns to scale, a growing literature finds empirical evidence against it, and large variation in industry-level returns to scale have been documented (Chirinko and Fazzari (1994), Basu and Fernald (1997), Gao and Kehrig (2016), Lafortune et al. (2021)). We find that Chinese firms in 2005 are on average constant returns to scale but there is large variation across industries.

The remainder of the paper is organized as follows. We introduce the data set in Section 2 and offers a brief explanation of the model and identification used by ZX in Section 3. We also talk about the theoretical consequences of using incorrect production elasticities and incorrect demand elasticities in Section 3. Section 4 presents the results and Section 5 concludes. Appendix provides derivations of the results in Section 3 and compare the differences between HK's data and our data.

2 Data

Our data source is the Chinese Annual Survey Data of Industries in 2005 collected by the National Bureau of Statistics of China.[1] This survey data has been used by previous studies including HK, Song et al. (2011), and David and Venkateswaran (2019), and ZX. It contains non-state firms with more than 5 million RMB (about $\$600,000$) in revenue and all the state-owned enterprises (SOEs).

The dataset is the same as the one used in ZX and contains rich information on firm-level value-added, wage expenditure, net value of fixed assets, sales, and costs. We clean the data and construct the depreciated real capital in the same way as in ZX who follows Brandt et al. (2012). The labor shares are corrected also in the same way as in ZX to account for the unobserved non-wage part. Table 1 displays the summary statistics of the data.

TABLE 1: Summary Statistics of Cleaned Data (2005)

Statistic	N	Mean	Min	Pctl(25)	Median	Pctl(75)	Max
value added	229,061	13,814.46	122	2,517	5,377	13,250	277,908
K	229,061	16,366.41	83.76	1,620.23	4,211.66	12,151.88	515,954.20
wL	229,061	2,730.73	80	583	1,188	2,665	78,956
revenue	229,061	50,184.74	2	9,500	19,457	45,994	11,041,153
cost	229,061	43,075.61	1	7,935	16,481	39,072	10,757,115
profits	229,061	2,370.47	−292,087	72	480	1,815	415,879
revenue/cost	229,061	1.21	0.81	1.08	1.14	1.25	4.68
wL/value added	229,061	0.32	0.01	0.12	0.23	0.42	3.15
wL^c/value added	229,061	0.850	0.033	0.310	0.621	1.108	8.338

Notes: wL^c is the corrected labor share. N is the number of firms.

[1] We acquire the data through Peking University.

3 Model

We use the model and identification method of ZX to estimate demand elasticities and production elasticities. In this section, we provide a brief explanation of the model setup and the identification method and refer readers to ZX for further details. We also demonstrate the predicted TFP gains and the reallocation of capital and labor across nests as well as the consequences of using incorrect production elasticities and demand elasticities.

3.1 Setup

In this economy, there are S industries, s_1, s_2, $\ldots$,s_S. An industry is referred to as s when its name is not specified. Firms, labeled as i, inside an industry have the same production elasticities α_s^K and α_s^L but differ in their Hicks-neutral productivity, A_i. Firm i produces product Y_i according to a Cobb-Douglas production function using capital K_i and labor L_i:

$$Y_i = A_i K_i^{\alpha_s^K} L_i^{\alpha_s^L}$$

Demand has nested constant elasticities of substitution (nested CES). Each nest, denoted by g, has a nest-specific demand elasticity, ϵ_g, which is larger than 1. The elasticity of substitution across nests is one. So $Y = \prod_{s \in \{s_1, \cdots, s_S\}} \prod_{g \in \{\bar{\mathsf{g}}(s), \underline{\mathsf{g}}(s)\}} Y_g^{\beta_g}$ and $Y_g = \left(\sum_{i \in \mathcal{G}(g)} Y_i^{\frac{\epsilon_g - 1}{\epsilon_g}} \right)^{\frac{\epsilon_g}{\epsilon_g - 1}}$. Y represents the aggregate production of all the firms, and Y_g is the production of nest g. An industry s may have two nests, one high-demand-elasticity nest $\bar{\mathsf{g}}(s)$ and one low-demand-elasticity nest $\underline{\mathsf{g}}(s)$, reflecting the different scope of product differentiation among firms producing similar products. It is also possible that an industry has only one nest, then $\bar{\mathsf{g}}(s) = \underline{\mathsf{g}}(s)$.

Firm i's profits Π_i are revenues minus the production costs:

$$\Pi_i = P_i Y_i - (R(1 + \tau_i^K) K_i + w(1 + \tau_i^L) L_i) e^{\delta_i}$$

The input distortions τ_i^L and τ_i^K enter the profits function as wedges multiplied with the market prices of capital and labor. The production costs contain unexpected cost shocks δ_i that are realized after firms choose capital, labor, and prices. The cost shocks δ_i follow a nest-specific normal distribution $\mathcal{N}\left(-\frac{\sigma_g^2}{2}, \sigma_g\right)$. This ensures that $\mathbb{E}[e^{\delta_i}] = 1$ is normalized to 1.

Firm i chooses its capital and labor and set its price to maximize the expected profits:

$$\begin{aligned}\max_{K_i,L_i,P_i} \mathbb{E}[\Pi_i] &= P_iY_i - (R(1+\tau_i^K)K_i + w(1+\tau_i^L)L_i) \\ \text{s.t. } Y_i &= A_iK_i^{\alpha_s^K}L_i^{\alpha_s^L} \\ &\quad \text{the nested CES demand}\end{aligned}$$

The first order conditions of the profit maximization give the relation between factor shares and model primitives such as input distortions. From the first order conditions, we can see that firm-specific input distortions, τ_i^L and τ_i^K, are the gaps between observed factor shares, $\frac{wL_i}{P_iY_i}$ and $\frac{RK_i}{P_iY_i}$, and the predicted factor shares under no input distortions, $\alpha_s^L\frac{\epsilon_g-1}{\epsilon_g}$ and $\alpha_s^K\frac{\epsilon_g-1}{\epsilon_g}$:

$$\underbrace{\ln(1+\tau_i^L)}_{\text{input distortions}} = \underbrace{\ln\left(\alpha_s^L\frac{\epsilon_g-1}{\epsilon_g}\right)}_{\substack{\text{predicted shares}\\ \text{w/o input distortions}}} - \underbrace{\ln\left(\frac{wL_i}{P_iY_i}\right)}_{\text{observed shares}}$$

$$\underbrace{\ln(1+\tau_i^K)}_{\text{input distortions}} = \underbrace{\ln\left(\alpha_s^K\frac{\epsilon_g-1}{\epsilon_g}\right)}_{\substack{\text{predicted shares}\\ \text{w/o input distortions}}} - \underbrace{\ln\left(\frac{RK_i}{P_iY_i}\right)}_{\text{observed shares}}$$

The unobserved cost shocks are not in the observed capital shares and labor shares. Since the factor shares have to be positive, input distortions τ_i^K and τ_i^L have to be larger than -1 by construction.

3.2 Reallocation and predicted TFP gains

The aggregate TFP gains can be decomposed into two parts: gains from reallocation within nests and gains from reallocation across nests.[2]

$$\text{TFP gains} = \frac{Y^*}{Y} = \prod_{s\in\{s_1,\cdots,s_S\}}\prod_{g\in\{\bar{\mathrm{g}}(s),\underline{\mathrm{g}}(s)\}} \underbrace{\left[\frac{\text{TFP}_g^*}{\text{TFP}_g}\right]^{\beta_g}}_{\text{gains within nests}} \cdot \underbrace{\left[\left(\frac{L_g^*}{L_g}\right)^{\alpha_g^L}\left(\frac{K_g^*}{K_g}\right)^{\alpha_g^K}\right]^{\beta_g}}_{\text{gains across nests}} \tag{1}$$

Y^* is the aggregate output when the input distortions are removed, i.e. $\tau_i^K = \tau_i^L = 0$. The same as in HK, nest g's TFP, TFP_g, is defined as $\frac{Y_g}{K_g^{\alpha_g^K}L_g^{\alpha_g^L}}$, where L_g and K_g are labor and

[2]HK only has the within-nest reallocation, because equalizing TFPR leads to no capital and labor flow across nests as explained by HK.

capital used in nest g. When input distortions are removed, nest g's TFP is denoted by TFP_g^*. We denote the capital and labor used in g without input distortions as L_g^* and K_g^*. The supply of aggregate capital and labor, K and L, is fixed. The wage and the capital rental rate clear the markets of capital and labor so that $\sum_{s\in\{s_1,\cdots,s_S\}}\sum_{g\in\{\bar{\mathrm{g}}(s),\underline{\mathrm{g}}(s)\}}\sum_{i\in\mathcal{G}(g)} K_i = \sum_{s\in\{s_1,\cdots,s_S\}}\sum_{g\in\{\bar{\mathrm{g}}(s),\underline{\mathrm{g}}(s)\}}\sum_{i\in\mathcal{G}(g)} K_i^* = K$ and $\sum_{s\in\{s_1,\cdots,s_S\}}\sum_{g\in\{\bar{\mathrm{g}}(s),\underline{\mathrm{g}}(s)\}}\sum_{i\in\mathcal{G}(g)} L_i = \sum_{s\in\{s_1,\cdots,s_S\}}\sum_{g\in\{\bar{\mathrm{g}}(s),\underline{\mathrm{g}}(s)\}}\sum_{i\in\mathcal{G}(g)} L_i^* = L$.

We follow HK to define the firm-level and nest-level revenue-based TFP, i.e. TFPR, as:

$$\begin{aligned}\text{TFPR}_i \equiv& P_i A_i = \frac{P_i Y_i}{K_i^{\alpha_s^K} L_i^{\alpha_s^L}} \\ \text{TFPR}_g \equiv& \frac{\sum_{i\in g} P_i Y_i}{K_g^{\alpha_s^K} L_g^{\alpha_s^L}}\end{aligned}$$

Then the nest-level TFP is:

$$\text{TFP}_g = \left(\sum_{i\in g}\left(A_i \cdot \frac{\text{TFPR}_g}{\text{TFPR}_i}\right)^{\epsilon_g - 1}\right)^{\frac{1}{\epsilon_g - 1}}$$

We follow HK and calculate A_i using $\frac{(P_i Y_i)^{\epsilon_g/(\epsilon_g-1)}}{K_i^{\alpha_s^K}(wL_i)^{\alpha_s^L}}$. However, the ratio $\text{TFPR}_i/\text{TFPR}_g$ and consequently TFP_g and TFP_g^* are different from HK due to non-constant returns to scale.

$$\frac{\text{TFPR}_i}{\text{TFPR}_g} = \underbrace{(1+\tau_i^K)^{\alpha_s^K}(1+\tau_i^L)^{\alpha_s^L}\left(\sum_{i\in\mathcal{G}(g)}\frac{P_i Y_i}{P_g Y_g (1+\tau_i^K)}\right)^{\alpha_s^K}\left(\sum_{i\in\mathcal{G}(g)}\frac{P_i Y_i}{P_g Y_g (1+\tau_i^L)}\right)^{\alpha_s^L}}_{\text{Same as CRS}} \cdot \left(\frac{P_i Y_i}{P_g Y_g}\right)^{1-\alpha_s^K-\alpha_s^L}$$

The last term in $\text{TFPR}_i/\text{TFPR}_g$ disappears under constant returns to scale (CRS). The rest is the same as the one in HK after replacing our notation of distortions by theirs. In equilibrium, firms' market shares in nests are (derivations are in Appendix A):

$$\frac{P_i Y_i}{P_g Y_g} = \frac{W_i}{\sum_{j\in g} W_j}$$

where,

$$W_i \equiv \left(\frac{1}{A_i}\right)^{\frac{1-\epsilon_g}{(1-\epsilon_g)(\alpha_s^L+\alpha_s^K)+\epsilon_g}}(1+\tau_i^K)^{\frac{\alpha^K(1-\epsilon_g)}{(1-\epsilon_g)(\alpha_s^L+\alpha_s^K)+\epsilon_g}}(1+\tau_i^L)^{\frac{\alpha^L(1-\epsilon_g)}{(1-\epsilon_g)(\alpha_s^L+\alpha_s^K)+\epsilon_g}}$$

Therefore, the the ratio $\text{TFPR}_i/\text{TFPR}_g$ is:

$$\begin{aligned}\frac{\text{TFPR}_i}{\text{TFPR}_g} =&\Gamma_i \cdot \left(\frac{\Gamma_i}{A_i}\right)^{\theta_g(1-\alpha_g^K-\alpha_g^L)} \cdot \left(\sum_{i\in\mathcal{G}(g)}\left(\frac{\Gamma_i}{A_i}\right)^{\theta_g}\right)^{-1} \\ &\cdot \left(\sum_{i\in\mathcal{G}(g)}\left(\frac{\Gamma_i}{A_i}\right)^{\theta_g}\frac{1}{1+\tau_i^K}\right)^{\alpha_g^K} \cdot \left(\sum_{i\in\mathcal{G}(g)}\left(\frac{\Gamma_i}{A_i}\right)^{\theta_g}\frac{1}{1+\tau_i^L}\right)^{\alpha_g^L}\end{aligned}$$

where

$$\begin{aligned}\Gamma_i \equiv&(1+\tau_i^K)^{\alpha_g^K}(1+\tau_i^L)^{\alpha_g^L} \\ \theta_g \equiv&\frac{1-\epsilon_g}{(1-\epsilon_g)(\alpha_g^K+\alpha_g^L)+\epsilon_g}\end{aligned}$$

Γ_i is a compound measure of firm i's capital and labor distortions based on its technology α_g^K and α_g^L. Then nest g's TFP_g is:

$$\text{TFP}_g = \left(\sum_{i\in\mathcal{G}(g)}\left(\frac{\Gamma_i}{A_i}\right)^{\theta_g}\right)^{\frac{\epsilon_g}{\epsilon_g-1}} \cdot \left(\sum_{i\in\mathcal{G}(g)}\left(\frac{\Gamma_i}{A_i}\right)^{\theta_g}\frac{1}{1+\tau_i^K}\right)^{-\alpha_g^K} \cdot \left(\sum_{i\in\mathcal{G}(g)}\left(\frac{\Gamma_i}{A_i}\right)^{\theta_g}\frac{1}{1+\tau_i^L}\right)^{-\alpha_g^L}$$

The derivations of $\text{TFPR}_i/\text{TFPR}_g$, and TFP_g are in Appendix A.

We denote all the variables under the scenario of no input distortions, i.e. $\tau_i^K = 0$ and $\tau_i^L = 0$, with a superscript $*$. The no-input-distortion equilibrium market share of firm i is:

$$\frac{P_i^* Y_i^*}{P_g^* Y_g^*} = \frac{A_i^{-\theta_g}}{\sum_{i\in\mathcal{G}(g)} A_i^{-\theta_g}}$$

and nest g's TFP under no input distortions is:

$$\text{TFP}_g^* = \left(\sum_{i\in\mathcal{G}(g)} A_i^{-\theta_g}\right)^{-\frac{1}{\theta_g}}$$

Different from HK, TFPR_i under no distortions is not equalized within a nest unless $\alpha_s^K + \alpha_s^L = 1$:

$$\frac{\text{TFPR}_i^*}{\text{TFPR}_g^*} = \left(\frac{P_i^* Y_i^*}{P_g^* Y_g^*}\right)^{1-\alpha_s^K-\alpha_s^L} = \left(\frac{A_i^{-\theta_g}}{\sum_{i\in\mathcal{G}(g)} A_i^{-\theta_g}}\right)^{1-\alpha_s^K-\alpha_s^L}$$

If $\alpha_s^K + \alpha_s^L = 1$, $\frac{\text{TFPR}_i^*}{\text{TFPR}_g^*} = 1$, so TFPR_i^* is equalized within g. The variation in TFPR_i^* in this case no longer indicates input distortions as in HK but reflects differences in firms' productivity, as argued by Haltiwanger et al. (2018).

The predicted TFP gains in nest g, i.e. changes in nest g's TFP_g when removing the input distortions in it, are the reciprocal of $\frac{\text{TFP}_g}{\text{TFP}_g^*}$:

$$\frac{\text{TFP}_g}{\text{TFP}_g^*} = \left(\frac{\sum_{i \in \mathcal{G}(g)} \left(\frac{\Gamma_i}{A_i}\right)^{\theta_g}}{\sum_{i \in \mathcal{G}(g)} A_i^{-\theta_g}} \right)^{\frac{\epsilon_g}{\epsilon_g - 1}} \cdot \left(\frac{\sum_{i \in \mathcal{G}(g)} \left(\frac{\Gamma_i}{A_i}\right)^{\theta_g} \frac{1}{1+\tau_i^K}}{\sum_{i \in \mathcal{G}(g)} A_i^{-\theta_g}} \right)^{-\alpha_g^K} \cdot \left(\frac{\sum_{i \in \mathcal{G}(g)} \left(\frac{\Gamma_i}{A_i}\right)^{\theta_g} \frac{1}{1+\tau_i^L}}{\sum_{i \in \mathcal{G}(g)} A_i^{-\theta_g}} \right)^{-\alpha_g^L} \tag{2}$$

The gains across nests are calculated using the ratios between nest-level labor and capital usage before and after the reallocation, as shown in Equation (1). The ratios can be written as:

$$\begin{aligned} \frac{L_g^*}{L_g} &= \frac{w^* L_g^* / (w^* L)}{w L_g / (wL)} \\ \frac{K_g^*}{K_g} &= \frac{K_g^* / K}{K_g / K} \end{aligned}$$

Since $wL_g/(wL)$ and K_g/K are directly observed, we only need to calculate $w^* L_g^*/(w^* L)$ and K_g^*/K:

$$\begin{aligned} \frac{w^* L_g^*}{w^* L} &= \frac{\beta_g \cdot \frac{\alpha_s^L}{\epsilon_g/(\epsilon_g - 1)\mathbb{E}[e^{\delta_i}]}}{\sum_g \beta_g \cdot \frac{\alpha_s^L}{\epsilon_g/(\epsilon_g - 1)\mathbb{E}[e^{\delta_i}]}} \\ \frac{K_g^*}{K} &= \frac{\beta_g \cdot \frac{\alpha_s^K}{\epsilon_g/(\epsilon_g - 1)\mathbb{E}[e^{\delta_i}]}}{\sum_g \beta_g \cdot \frac{\alpha_s^K}{\epsilon_g/(\epsilon_g - 1)\mathbb{E}[e^{\delta_i}]}} \end{aligned}$$

Using the formulas above, aggregate TFP gains can be calculated once the parameters, α_g^L, α_g^K, ϵ_g, and β_g, are identified. β_g is simply the expenditure share of nest g. The estimation of the remaining parameters will be discussed in Section **??**.

3.3 Consequences of calibrating α_g^K and α_g^L using a benchmark economy

This section explains why calibrating Chinese α_g^K and α_g^L using a benchmark economy does not affect the predicted TFP gains through the estimated values of input distortions but does affect the predicted gains through the aggregation and through the estimated firm productivity A_i. Denote the production elasticities of the benchmark economy as $\tilde{\alpha}_g^K$ and $\tilde{\alpha}_g^L$ and define the difference between the production elasticities of the benchmark economy and those of China as:

$$\begin{aligned}\delta_g^K &\equiv \frac{\tilde{\alpha}_g^K}{\alpha_g^K}\\ \delta_g^L &\equiv \frac{\tilde{\alpha}_g^K}{\alpha_g^L}\end{aligned}$$

δ^K and δ^L can also be interpreted as measurement errors of the production elasticities. The input distortions estimated using the calibrated $\tilde{\alpha}_g^K$ and $\tilde{\alpha}_g^L$ are denoted as $\tilde{\tau}_i^K$ and $\tilde{\tau}_i^L$:

$$\begin{aligned}\ln(1+\tilde{\tau}_i^K) &= \ln(1+\tau_i^K) + \ln(\delta_g^K)\\ \ln(1+\tilde{\tau}_i^L) &= \ln(1+\tau_i^L) + \ln(\delta_g^L)\end{aligned}$$

τ_i^K and τ_i^L are the input distortions estimated using Chinese α_g^K and α_g^L.

Using $\tilde{\alpha}_g^K$ and $\tilde{\alpha}_g^L$, firms' productivity $\tilde{A}_i$ is measures as:

$$\tilde{A}_i = \frac{(P_iY_i)^{\epsilon_g/(\epsilon_g-1)}}{K_i^{\tilde{\alpha}_g^K}(wL_i)^{\tilde{\alpha}_g^L}}$$

Nest g's TFP using $\tilde{\alpha}_g^K$ and $\tilde{\alpha}_g^L$ is:

$$\begin{aligned}\widetilde{\mathrm{TFP}}_g = &\left(\sum_{i\in\mathcal{G}(g)}\left(\frac{\tilde{\Gamma}_i}{\tilde{A}_i}\right)^{\tilde{\theta}_g}\right)^{\frac{\epsilon_g}{\epsilon_g-1}}\cdot\\ &\left(\sum_{i\in\mathcal{G}(g)}\left(\frac{\tilde{\Gamma}_i}{\tilde{A}_i}\right)^{\tilde{\theta}_g}\frac{1}{(1+\tau_i^K)\delta_g^K}\right)^{-\tilde{\alpha}_g^K}\cdot\left(\sum_{i\in\mathcal{G}(g)}\left(\frac{\tilde{\Gamma}_i}{\tilde{A}_i}\right)^{\tilde{\theta}_g}\frac{1}{(1+\tau_i^L)\delta_g^L}\right)^{-\tilde{\alpha}_g^L}\end{aligned}$$

where $\tilde{\Gamma}_i = (1+\tilde{\tau}_i^K)^{\tilde{\alpha}_g^K}(1+\tilde{\tau}_i^L)^{\tilde{\alpha}_g^L} = (1+\tau_i^K)^{\tilde{\alpha}_g^K}(1+\tau_i^L)^{\tilde{\alpha}_g^L}(\delta_g^K)^{\tilde{\alpha}_g^K}(\delta_g^L)^{\tilde{\alpha}_g^L}$ and $\tilde{\theta}_g = \frac{1-\epsilon_g}{(1-\epsilon_g)(\tilde{\alpha}_g^K+\tilde{\alpha}_g^L)+\epsilon_g}$.

It turns out that δ_g^L and δ_g^K cancel out in $\widetilde{\text{TFP}}_g$:

$$\widetilde{\text{TFP}}_g = \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{(1+\tau_i^K)^{\tilde{\alpha}_g^K}(1+\tau_i^L)^{\tilde{\alpha}_g^L}}{\tilde{A}_i}\right)^{\tilde{\theta}_g}\right)^{\frac{\epsilon_g}{\epsilon_g-1}}$$
$$\cdot \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{(1+\tau_i^K)^{\tilde{\alpha}_g^K}(1+\tau_i^L)^{\tilde{\alpha}_g^L}}{\tilde{A}_i}\right)^{\tilde{\theta}_g}\frac{1}{(1+\tau_i^K)}\right)^{-\tilde{\alpha}_g^K}$$
$$\cdot \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{(1+\tau_i^K)^{\tilde{\alpha}_g^K}(1+\tau_i^L)^{\tilde{\alpha}_g^L}}{\tilde{A}_i}\right)^{\tilde{\theta}_g}\frac{1}{(1+\tau_i^L)}\right)^{-\tilde{\alpha}_g^L}$$

When there is no input distortions:

$$\widetilde{\text{TFP}}_g^* = \sum_{i\in\mathcal{G}(g)} \left(\tilde{A}_i^{-\tilde{\theta}_g}\right)^{-\frac{1}{\tilde{\theta}_g}}$$

Predicted TFP gains of nest g are the inverse of:

$$\begin{aligned}\frac{\widetilde{\text{TFP}}_g}{\widetilde{\text{TFP}}_g^*} =& \frac{1}{\left(\sum_{i\in\mathcal{G}(g)}\tilde{A}_i^{-\tilde{\theta}_g}\right)^{-\frac{1}{\tilde{\theta}_g}}} \cdot \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{(1+\tau_i^K)^{\tilde{\alpha}_g^K}(1+\tau_i^L)^{\tilde{\alpha}_g^L}}{\tilde{A}_i}\right)^{\tilde{\theta}_g}\right)^{\frac{\epsilon_g}{\epsilon_g-1}} \\
&\cdot \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{(1+\tau_i^K)^{\tilde{\alpha}_g^K}(1+\tau_i^L)^{\tilde{\alpha}_g^L}}{\tilde{A}_i}\right)^{\tilde{\theta}_g}\frac{1}{(1+\tau_i^K)}\right)^{-\tilde{\alpha}_g^K} \\
&\cdot \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{(1+\tau_i^K)^{\tilde{\alpha}_g^K}(1+\tau_i^L)^{\tilde{\alpha}_g^L}}{\tilde{A}_i}\right)^{\tilde{\theta}_g}\frac{1}{(1+\tau_i^L)}\right)^{-\tilde{\alpha}_g^L} \\
=& \left(\frac{\sum_{i\in\mathcal{G}(g)}\left(\frac{(1+\tau_i^K)^{\tilde{\alpha}_g^K}(1+\tau_i^L)^{\tilde{\alpha}_g^L}}{\tilde{A}_i}\right)^{\tilde{\theta}_g}}{\sum_{i\in\mathcal{G}(g)}\tilde{A}_i^{-\tilde{\theta}_g}}\right)^{\frac{\epsilon_g}{\epsilon_g-1}} \\
&\cdot \left(\frac{\sum_{i\in\mathcal{G}(g)}\left(\frac{(1+\tau_i^K)^{\tilde{\alpha}_g^K}(1+\tau_i^L)^{\tilde{\alpha}_g^L}}{\tilde{A}_i}\right)^{\tilde{\theta}_g}\frac{1}{(1+\tau_i^K)}}{\sum_{i\in\mathcal{G}(g)}\tilde{A}_i^{-\tilde{\theta}_g}}\right)^{-\tilde{\alpha}_g^K} \cdot \left(\frac{\sum_{i\in\mathcal{G}(g)}\left(\frac{(1+\tau_i^K)^{\tilde{\alpha}_g^K}(1+\tau_i^L)^{\tilde{\alpha}_g^L}}{\tilde{A}_i}\right)^{\tilde{\theta}_g}\frac{1}{(1+\tau_i^L)}}{\sum_{i\in\mathcal{G}(g)}\tilde{A}_i^{-\tilde{\theta}_g}}\right)^{-\tilde{\alpha}_g^L}\end{aligned} \tag{3}$$

The second equation is simply rearranging $\frac{1}{\left(\sum_{i\in\mathcal{G}(g)}\tilde{A}_i^{-\tilde{\theta}_g}\right)^{-\frac{1}{\tilde{\theta}_g}}}$.

The predicted TFP gains of nest g using Chinese α_g^K and α_g^L are given in Equation (2). To make the comparison with Equation (3) easier, we rewrite Equation (2) by plugging in the formula of Γ_i:

$$\frac{\text{TFP}_g}{\text{TFP}_g^*} = \left(\frac{\sum_{i\in\mathcal{G}(g)}\left(\frac{(1+\tau_i^K)^{\alpha_g^K}(1+\tau_i^L)^{\alpha_g^L}}{A_i}\right)^{\theta_g}}{\sum_{i\in\mathcal{G}(g)}A_i^{-\theta_g}}\right)^{\frac{\epsilon_g}{\epsilon_g-1}} \tag{4}$$

$$\cdot\left(\frac{\sum_{i\in\mathcal{G}(g)}\left(\frac{(1+\tau_i^K)^{\alpha_g^K}(1+\tau_i^L)^{\alpha_g^L}}{A_i}\right)^{\theta_g}\frac{1}{1+\tau_i^K}}{\sum_{i\in\mathcal{G}(g)}A_i^{-\theta_g}}\right)^{-\alpha_g^K}\cdot\left(\frac{\sum_{i\in\mathcal{G}(g)}\left(\frac{(1+\tau_i^K)^{\alpha_g^K}(1+\tau_i^L)^{\alpha_g^L}}{A_i}\right)^{\theta_g}\frac{1}{1+\tau_i^L}}{\sum_{i\in\mathcal{G}(g)}A_i^{-\theta_g}}\right)^{-\alpha_g^L}$$

Comparing Equation (3) and (4), we can see that the biases of the predicted TFP gains caused by calibrating the production elasticities are not due to the biases in the estimated input distortions but how the input distortions are aggregated within firms, i.e. the $\tilde{\alpha}_g^K$ and $\tilde{\alpha}_g^L$ in $(1+\tau_i^K)^{\tilde{\alpha}_g^K}(1+\tau_i^L)^{\tilde{\alpha}_g^L}$, and across firms, i.e. the $\tilde{\alpha}_g^K$ and $\tilde{\alpha}_g^L$ in the indexes of the second line of Equation (3) and $\tilde{\theta}_g$, and through the estimated $\tilde{A}_i$.

3.4 Consequences of using homogeneous demand elasticities

In this section, we talk about the consequence of using the correct production elasticities but incorrect homogeneous demand elasticities. We denote ϵ_g as the true demand elasticities and $\tilde{\epsilon}$ as the homogeneous demand elasticity assumed by researchers. We define δ_g^ϵas:

$$\ln(\delta_g^\epsilon) = \ln\left(\frac{\tilde{\epsilon}}{\tilde{\epsilon}-1}\right) - \ln\left(\frac{\epsilon_g}{\epsilon_g-1}\right)$$

Then the input distortions, $\tilde{\tau}_i^{\epsilon,L}$ and $\tilde{\tau}_i^{\epsilon,K}$, estimated using $\tilde{\epsilon}$ are:

$$\ln(1+\tilde{\tau}_i^{\epsilon,L}) = \ln(1+\tau_i^L) + \ln(\delta_g^\epsilon)$$
$$\ln(1+\tilde{\tau}_i^{\epsilon,K}) = \ln(1+\tau_i^K) + \ln(\delta_g^\epsilon)$$

where τ_i^L and τ_i^K are the input distortions estimated using ϵ_g.

Following the same procedure as in Section 3.3, we can get the TFP_g of nest g using $\tilde{\epsilon}$, which is denoted as $\widetilde{\mathrm{TFP}}_g^{\epsilon}$:

$$\widetilde{\mathrm{TFP}}_g^{\epsilon} = \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{(1+\tau_i^K)^{\alpha_g^K}(1+\tau_i^L)^{\alpha_g^L}}{\tilde{A}_i^{\epsilon}}\right)^{\tilde{\theta}_g^{\epsilon}}\right)^{\frac{\tilde{\epsilon}}{\tilde{\epsilon}-1}}$$
$$\cdot \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{(1+\tau_i^K)^{\alpha_g^K}(1+\tau_i^L)^{\alpha_g^L}}{\tilde{A}_i^{\epsilon}}\right)^{\tilde{\theta}_g^{\epsilon}} \frac{1}{(1+\tau_i^K)}\right)^{-\alpha_g^K}$$
$$\cdot \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{(1+\tau_i^K)^{\alpha_g^K}(1+\tau_i^L)^{\alpha_g^L}}{\tilde{A}_i^{\epsilon}}\right)^{\tilde{\theta}_g^{\epsilon}} \frac{1}{(1+\tau_i^L)}\right)^{-\alpha_g^L}$$

where $\tilde{A}_i^{\epsilon} = \frac{(P_iY_i)^{\tilde{\epsilon}/(\tilde{\epsilon}-1)}}{K_i^{\alpha_g^K}(wL_i)^{\alpha_g^L}}$, $\tilde{\theta}_g^{\epsilon} = \frac{1-\tilde{\epsilon}}{(1-\tilde{\epsilon})(\alpha_g^K+\alpha_g^L)+\tilde{\epsilon}}$.

When there is no input distortions, the TFP of nest g is denoted as $\widetilde{\mathrm{TFP}}_g^{\epsilon *}$:

$$\widetilde{\mathrm{TFP}}_g^{\epsilon *} = \sum_{i\in\mathcal{G}(g)} \left((\tilde{A}_i^{\epsilon})^{-\tilde{\theta}_g^{\epsilon}}\right)^{-\frac{1}{\tilde{\theta}_g^{\epsilon}}}$$

Then the TFP gains of nest g is:

$$\frac{\widetilde{\mathrm{TFP}}_g^{\epsilon}}{\widetilde{\mathrm{TFP}}_g^{\epsilon *}} = \left(\frac{\sum_{i\in\mathcal{G}(g)} \left(\frac{(1+\tau_i^K)^{\alpha_g^K}(1+\tau_i^L)^{\alpha_g^L}}{\tilde{A}_i^{\epsilon}}\right)^{\tilde{\theta}_g^{\epsilon}}}{\sum_{i\in\mathcal{G}(g)}(\tilde{A}_i^{\epsilon})^{-\tilde{\theta}_g^{\epsilon}}}\right)^{\frac{\tilde{\epsilon}}{\tilde{\epsilon}-1}} \tag{5}$$
$$\cdot \left(\frac{\sum_{i\in\mathcal{G}(g)} \left(\frac{(1+\tau_i^K)^{\alpha_g^K}(1+\tau_i^L)^{\alpha_g^L}}{\tilde{A}_i^{\epsilon}}\right)^{\tilde{\theta}_g^{\epsilon}} \frac{1}{1+\tau_i^K}}{\sum_{i\in\mathcal{G}(g)}(\tilde{A}_i^{\epsilon})^{-\tilde{\theta}_g^{\epsilon}}}\right)^{-\alpha_g^K} \cdot \left(\frac{\sum_{i\in\mathcal{G}(g)} \left(\frac{(1+\tau_i^K)^{\alpha_g^K}(1+\tau_i^L)^{\alpha_g^L}}{\tilde{A}_i^{\epsilon}}\right)^{\tilde{\theta}_g^{\epsilon}} \frac{1}{1+\tau_i^L}}{\sum_{i\in\mathcal{G}(g)}(\tilde{A}_i^{\epsilon})^{-\tilde{\theta}_g^{\epsilon}}}\right)^{-\alpha_g^L}$$

Comparing Equation (4) and (5) , we can see that the biases of the predicted TFP gains caused by assuming homogeneous demand elasticities are not due to the biases in the estimated input distortions but how the input distortions are aggregated across firms, i.e. the $\tilde{\theta}_g^{\epsilon}$ and $\tilde{\epsilon}$, and through the estimated $\tilde{A}_i^{\epsilon}$.

4 Results

All the demand and production estimates are the same as those in ZX. The demand elasticities and the latent nest structure are estimated by maximizing the likelihood of firm-level markups, which are measured using firm-level revenue-cost ratios. The industry-specific production elasticities are estimated by maximizing the likelihood of firm-level labor shares and capital shares. Further details about identification and estimation are in ZX.

We follow ZX to use revenue-cost ratios as measurement of firms' markups because firms' returns to scale are on average constant. We do not use methods developed in De Loecker and Warzynski (2012) because De Loecker and Warzynski (2012) relies on estimating production elasticities using the literature of estimating production function (Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg et al. (2015)) where some production input is required to be a monotone function of the unobserved productivity, but this monotonicity is unfortunately not guaranteed under firm-specific input distortions. In fact, it is usually believed that domestic private firms in China are more productive but more financially constrained than SOEs, which implies that sometimes more productive firms use less production inputs. Furthermore, we do not observe physical production like most studies. Markups estimated using De Loecker and Warzynski (2012) using nominal production can be uninformative about the true markups as argued by Bond et al. (2021).

In the rest of this section, we first report the summary statistics of the estimates and compare the production elasticities of American firms and those estimated using Chinese firm-level data. We then show the predicted TFP gains and how they respond to relaxing the assumptions imposed in HK.

4.1 Estimated parameters

Table 2 displays the distribution of industry-level firm counts. The first row is that of single-nest industries and the second row is that of two-nest industries. 462 industries are estimated as having two nest and 61 as having one nest. Industries containing two nests tend to contain more firms.

TABLE 2: Distribution of industry-level firm counts

	N	Mean	Min	Pctl(25)	Median	Pctl(75)	Max
One nest	61	23	2	6	15	27	237
Two nests	462	494	12	118	256	545	9,947

Table 3: Summary statistics of selected estimated parameters

	Mean	St. Dev.	Pctl(10)	Pctl(25)	Median	Pctl(75)	Pctl(90)
w_s	0.73	0.16	0.54	0.66	0.75	0.83	0.89
ϵ_g	8.49	3.26	3.99	6.50	8.57	10.27	12.85
ϵ_g (cost)[1]	9.37	3.51	4.50	7.44	9.20	10.91	14.16
ϵ_g (revenue)[2]	9.07	3.58	4.14	7.06	9.14	10.76	14.12
$\alpha_s^K + \alpha_s^L$	1.04	0.58	0.44	0.64	0.88	1.22	1.88
$\alpha_s^K + \alpha_s^L$(cost)[1]	0.96	0.55	0.36	0.53	0.82	1.16	1.83
$\alpha_s^K + \alpha_s^L$ (revenue)[2]	0.97	0.55	0.36	0.54	0.84	1.16	1.83

[1] The distribution is weighted by firms' costs.
[2] The distribution is weighted by firms' revenues.

The first row in Table 3 shows the ex-ante probability of belonging to the high demand-elasticity nest. In 90% of the 462 industries that contain two nests, it is more likely to be in the high demand-elasticity nest. This suggests that achieving a high level of demand elasticities is difficult. The second row in Table 3 is the distribution of demand elasticities across 229,064 firms. The third and forth row weight each firm by their costs and revenues respectively. Depending on whether firms are weighted by their costs or revenues, the average demand elasticity is between 8.5 and 9.4 and the median is between 8.6 and 9.2. There is large variation in demand elasticities with the top 10 percentile about three times larger than the bottom 10 percentile. The fifth till seventh rows report the distribution of estimated returns to scale across firms. The fifth row is the unweighted distribution, and the other two rows are weighted by costs and revenues respectively. On average, the industrial firms have constant returns to scale, but there is large variation across industries.

The demand elasticities estimated in ZX are on average 3 times larger than the 3 assumed in HK, but it seems that ZX's estimates are closer to other studies' estimates while HK's value is around or below the lower bound. A more detailed comparison with other studies' estimates is provided in ZX.

Figure 1 shows the distribution of production elasticities of American firms and the production elasticities estimated using Chinese firm-level survey data. It shows that American firms are more capital intensive than Chinese firms.

4.2 TFP gains and model assumptions

The predicted TFP gains using estimated demand elasticities and production elasticities are 45%. The TFP gains from reallocating capital and labor within nests are 32% and from reallocating across nests 9%.

Figure 1: α_K and α_L of Chinese firms and American firms

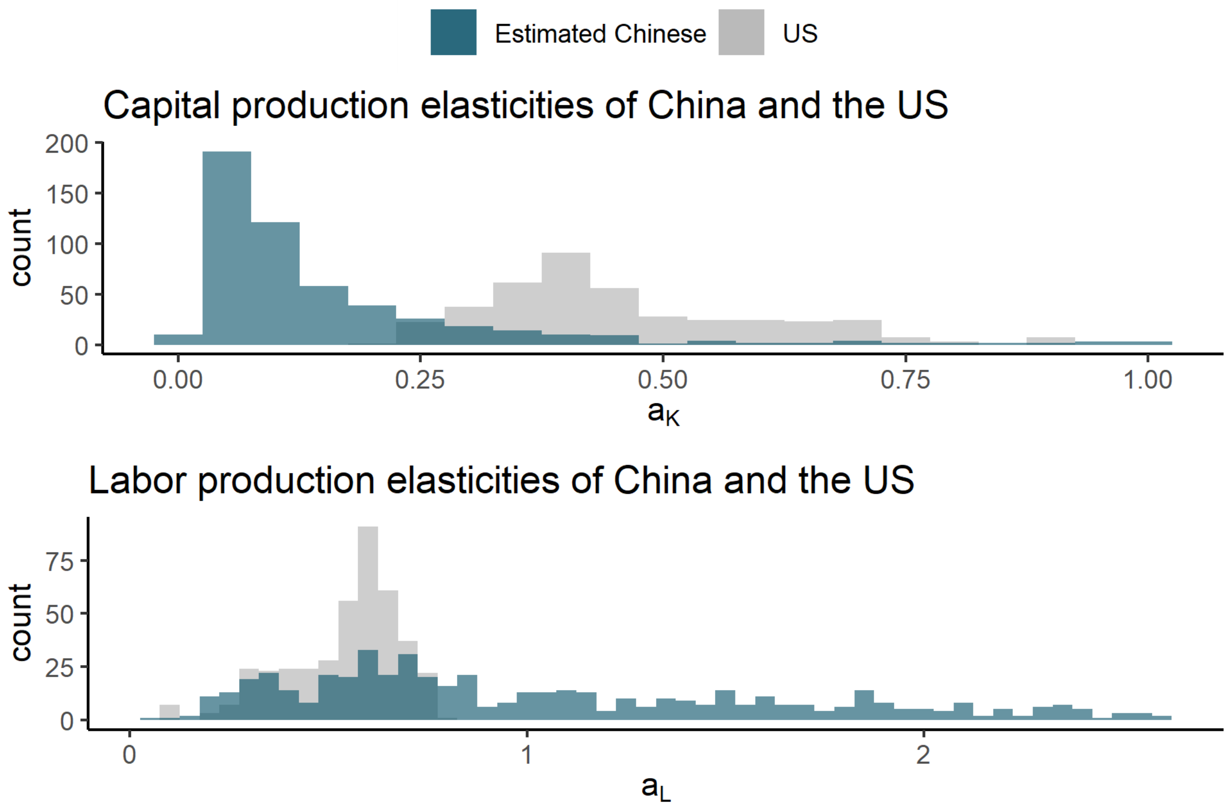


Table 4: TFP gains in China (2005)

within industry (%)	across industry (%)	total (%)
32.1	9.4	44.5

We then do an experiment by taking HK's parameters and then calculating the predicted TFP gains when replacing their estimates by ZX. There is some discrepancy between the datasets used by HK and by ZX, though both datasets are Chinese annual survey data. Appendix B compares the aggregate variables calculated using both datasets with the relevant macro variables published in the Chinese Yearbooks and find that our dataset is closer to the data used to construct the yearbooks. In Table 5, we do the same experiment using both datasets to offer a more comprehensive comparison to the value calculated in HK. When using HK's data, each industry can have only one nest because we do not observe firms' costs in their data. However, we can still estimate industry-specific demand elasticities using the method in ZX and replace HK's demand elasticities by estimated industry-specific demand elasticities. Therefore, we also include a scenario in the experiment on our data where each industry is a nest.

Comparing row 1 and 2 in the first half and the second halfof Table 5, we can see that changing demand elasticities from 3, the value assumed in HK, to 8.5, the average of our estimates, triples or quadruples the predicted TFP gains. Going from 8.5 to heterogeneous demand elasticities changes the predicted TFP gains by about or less than $\frac{1}{6}$ in most scenario except for the case of using HK's data and the estimated production elasticities α. The change in predicted TFP gains is less than $\frac{1}{10}$ when introducing two nests inside industries.

Table 5: Within-type TFP gains in China (2005) comparison across models

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-nest industries)[1]	299.5
HK	Our estimators	8.5	10.3
HK	Our estimators	heterogeneous (one-nest industries)[1]	6.1
Our	calibrated using US firms (HK)	3	233.8
Our	calibrated using US firms (HK)	8.5	739.4
Our	calibrated using US firms (HK)	heterogeneous (one-nest industries)[1]	628.2
Our	calibrated using US firms (HK)	heterogeneous (two-nest industries)[2]	628.5
Our	Our estimators	8.5	28
Our	Our estimators	heterogeneous (one-nest industries)[1]	29.4
Our	Our estimators	heterogeneous (two-nest industries)[2]	32.1

[1] Each industry contains one nest.
[2] Each industry can contain one or two nests.

Table 6 displays the reallocation across nests by showing how the nest-level labor and capital usage change after removing the input distortions. More than half of the nests reduce their capital and labor usage while some nests' capital and labor are 10 and 7 times larger.

Table 6: Changes in nest-level labor and capital

Statistic	Mean	Min	Pctl(25)	Median	Pctl(75)	Max
$\frac{L_g^*}{L_g}$	1.03	0.08	0.62	0.86	1.32	6.91
$\frac{K_g^*}{K_g}$	1.12	0.05	0.51	0.82	1.33	10.49

Changes in the nest-level input usage from removing the input distortions reflect the allocation across nests. Figure 2 show how these changes are related to nests' characteristics, namely production elasticities, α_s^K and α_s^L, and demand elasticities ϵ_g. There is no clear pattern between changes in input usage and demand elasticities as shown in Figure 2a and Figure 2b. However, Figure 2c and Figure 2d show that nests with higher demand for production inputs due to their technology, i.e. higher α_s^K and α_s^L, tend to receive more production inputs in the reallocation.

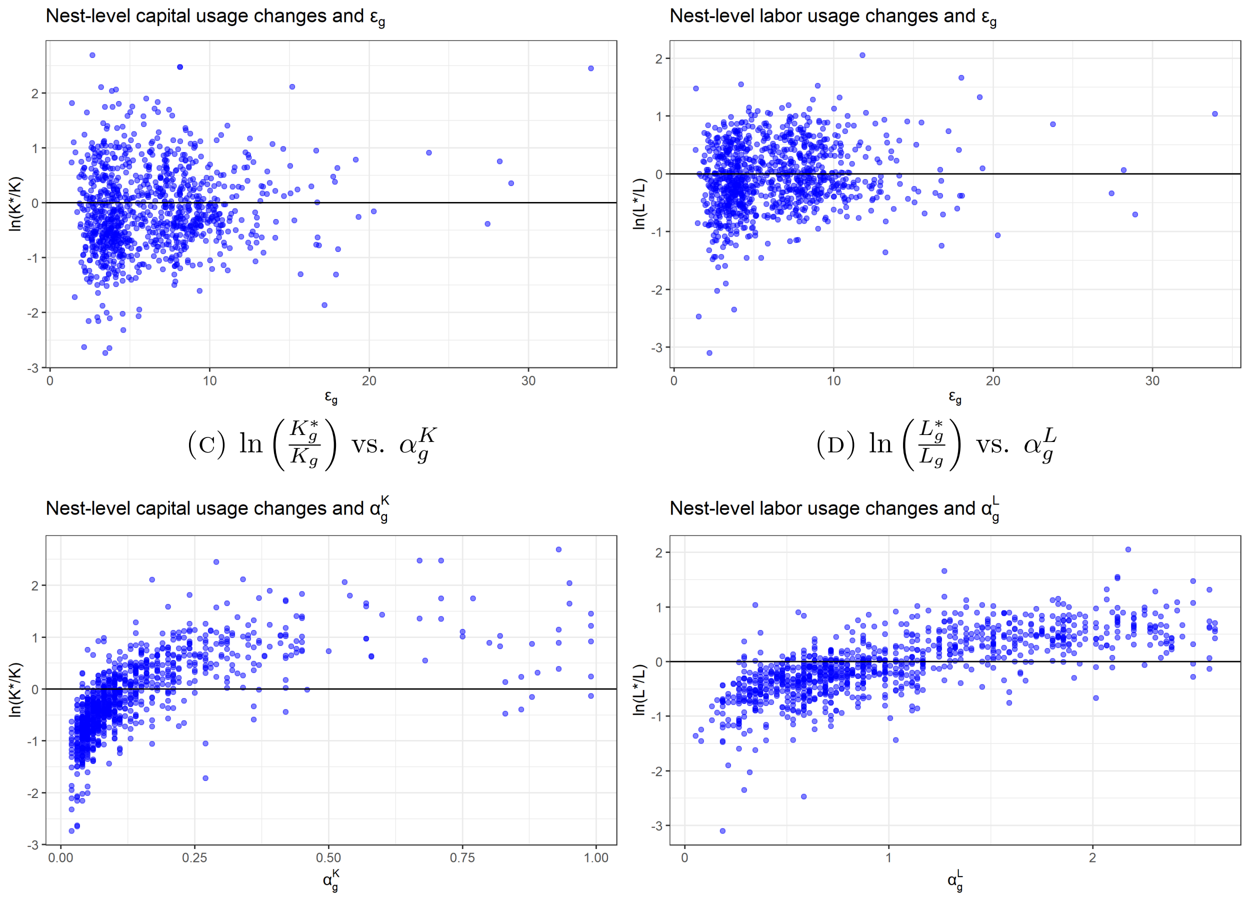
Figure 2: Changes in nest-level capital and labor usage

(A) $\ln\left(\frac{K_g^*}{K_g}\right)$ vs. ϵ_g

(B) $\ln\left(\frac{L_g^*}{L_g}\right)$ vs. ϵ_g

(C) $\ln\left(\frac{K_g^*}{K_g}\right)$ vs. α_g^K

(D) $\ln\left(\frac{L_g^*}{L_g}\right)$ vs. α_g^L



5 Conclusion

Measuring the TFP costs of misallocation due to input distortions has generated great interest, especially following Hsieh and Klenow (2009), but how to estimate demand and production parameters using firm-level data remains a challenge. Our paper uses the method and estimates in Zhang and Xia (2023) to review whether and how the predicted TFP gains in China from removing input distortion are sensitive to the assumption of a low level of homogeneous demand elasticities and that the Chinese firms' technology is the same as those in a benchmark economy, i.e. the US. Although in theory, these assumptions lead to contaminated estimation of input distortions and the predicted TFP gains as explained in Haltiwanger et al. (2018), our results find that the dispersion in demand elasticities has a much smaller impact than the average level of demand elasticities. Although the production elasticities between American and Chinese firms differ systematically, calibrating Chinese firms production function using American firms' parameter does not affect the dispersion in estimated input distortions and therefore does not affect the predicted TFP gains directly through the input distortions, but it affects the predicted gains through the aggregation across firms and through the estimated firm productivity.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015): "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83, 2411–2451.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020): "The Fall of the Labor Share and the Rise of Superstar Firms," *The Quarterly Journal of Economics*, 135, 645–709.

Basu, S. and J. G. Fernald (1997): "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy*, 105, 249–283.

Bond, S., A. Hashemi, G. Kaplan, and P. Zoch (2021): "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data," *Journal of Monetary Economics*, 121, 1–14.

Brandt, L., J. Van Biesebroeck, and Y. Zhang (2012): "Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing," *Journal of Development Economics*, 97, 339–351.

——— (2014): "Challenges of Working with the Chinese NBS Firm-Level Data," *China Economic Review*, 30, 339–352.

Chirinko, R. S. and S. M. Fazzari (1994): "Economic Fluctuations, Market Power, and Returns to Scale: Evidence from Firm-Level Data," *Journal of Applied Econometrics*, 9, 47–69.

David, J. M. and V. Venkateswaran (2019): "The Sources of Capital Misallocation," *American Economic Review*, 109, 2531–2567.

De Loecker, J. and F. Warzynski (2012): "Markups and Firm-Level Export Status," *American Economic Review*, 102, 2437–2471.

Gao, W. and M. Kehrig (2016): "Returns to Scale, Productivity and Competition: Empirical Evidence from U.S. Manufacturing and Construction Establishments," *SSRN Journal*.

Haltiwanger, J., R. Kulick, and C. Syverson (2018): "Misallocation Measures: The Distortion That Ate the Residual," Working Paper 24199, National Bureau of Economic Research.

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 46.

Lafortune, J., E. G. Lewis, J. Pablo Martínez, and J. Tessada (2021): "Changing Returns to Scale in Manufacturing 1880-1930: The Rise of (Skilled) Labor?" NBER working paper 28633.

Levinsohn, J. and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *The Review of Economic Studies*, 70, 317–341.

Melitz, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725.

Olley, G. S. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

Restuccia, D. and R. Rogerson (2008): "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments," *Review of Economic Dynamics*, 11, 707–720.

Ruzic, D. and S.-J. Ho (2021): "Returns to Scale, Productivity Measurement, and Trends in U.S. Manufacturing Misallocation," *The Review of Economics and Statistics*, 1–47.

Song, Z., K. Storesletten, and F. Zilibotti (2011): "Growing Like China," *American Economic Review*, 101, 196–233.

Zhang, X. and J. Xia (2023): "Factor Income Shares and Input Distortions in China," Working paper.

Appendix

A Derivation of predicted TFP gains

We first show how to derive the optimal prices. The optimal prices are always the expected marginal cost times $\epsilon_g/(\epsilon_g - 1)$. For some given Y_i, firms' profits maximization problem can be formulated as, :

$$\begin{aligned}&\min_{K_i,L_i}(R(1+\tau_i^K)K_i + w(1+\tau_i^L))\mathbb{E}[e^{\delta_i}]\\ &\text{s.t. } A_i K_i^{\alpha_s^K} L_i^{\alpha_s^L} \geq Y_i\end{aligned}$$

Expected marginal cost is the Lagrange multiplier of its Lagrange function

$$\min_{K_i,L_i}(R(1+\tau_i^K)K_i + w(1+\tau_i^L))\mathbb{E}[e^{\delta_i}] - \lambda(A_i K_i^{\alpha_s^K} L_i^{\alpha_s^L} - Y_i)$$

Solving it gives expected marginal cost:

$$\mathbb{E}[MC(Y_i)] = \left(\frac{1}{A_i}\right)^{\frac{1}{\alpha_g^L+\alpha_g^K}} Y_i^{\frac{1-\alpha_g^L-\alpha_g^K}{\alpha_g^L+\alpha_g^K}} \left(\frac{R(1+\tau_i^K)}{\alpha_g^K}\right)^{\frac{\alpha_g^K}{\alpha_g^L+\alpha_g^K}} \left(\frac{w(1+\tau_i^L)}{\alpha_g^L}\right)^{\frac{\alpha_g^L}{\alpha_g^L+\alpha_g^K}} \mathbb{E}[e^{\delta_i}|i \in \mathcal{G}(g)]$$

where $\alpha_g^K = \alpha_s^K$ and $\alpha_g^L = \alpha_s^L$ for firm i from nest g in industry s. The optimal prices are:

$$P_i = \frac{\epsilon_g}{\epsilon_g - 1} \cdot \underbrace{\left(\frac{1}{A_i}\right)^{\frac{1}{\alpha_g^L+\alpha_g^K}} Y_i^{\frac{1-\alpha_g^L-\alpha_g^K}{\alpha_g^L+\alpha_g^K}} \left(\frac{R(1+\tau_i^K)}{\alpha_g^K}\right)^{\frac{\alpha_g^K}{\alpha_g^L+\alpha_g^K}} \left(\frac{w(1+\tau_i^L)}{\alpha_g^L}\right)^{\frac{\alpha_g^L}{\alpha_g^L+\alpha_g^K}} \mathbb{E}[e^{\delta_i}]}_{\text{expected marginal cost}}$$

The nest-level TFP as a weighted sum of firm-level TFP is the same as the one in HK because the expression only requires the type-level aggregator to be CES:

$$
\begin{aligned}
\mathrm{TFP}_g =& \mathrm{TFPR}_g \cdot \frac{1}{P_g} \\
=& \mathrm{TFPR}_g \cdot \left(\sum_{i \in \mathcal{G}(g)} P_i^{1-\epsilon_g} \right)^{1/(\epsilon_g - 1))} \\
=& \mathrm{TFPR}_g \cdot \left(\sum_{i \in \mathcal{G}(g)} \left(\frac{A_i}{\mathrm{TFPR}_i} \right)^{\epsilon_g - 1} \right)^{1/(\epsilon_g - 1))} \\
=& \left(\sum_{i \in \mathcal{G}(g)} \left(A_i \cdot \frac{\mathrm{TFPR}_g}{\mathrm{TFPR}_i} \right)^{\epsilon_g - 1} \right)^{\frac{1}{\epsilon_g - 1}}
\end{aligned} \tag{6}
$$

From the definition of TFPR:

$$
\begin{aligned}
\mathrm{TFPR}_g &= \left(\frac{P_g Y_g}{K_g} \right)^{\alpha_g^K} \left(\frac{P_g Y_g}{L_g} \right)^{\alpha_g^L} (P_g Y_g)^{1-\alpha_s^K - \alpha_g^L} \\
\mathrm{TFPR}_i &= \left(\frac{P_i Y_i}{K_i} \right)^{\alpha_g^K} \left(\frac{P_i Y_i}{L_i} \right)^{\alpha_g^L} (P_i Y_i)^{1-\alpha_g^K - \alpha_g^L}
\end{aligned}
$$

Firms' profit maximization also gives:

$$
\begin{aligned}
\frac{K_i}{P_g Y_g} =& \frac{\epsilon_g - 1}{\epsilon_g} \cdot \frac{\alpha_g^K}{(1+\tau_i^K)R} \cdot \frac{P_i Y_i}{P_g Y_g} \\
\frac{L_i}{P_g Y_g} =& \frac{\epsilon_g - 1}{\epsilon_g} \cdot \frac{\alpha_g^L}{(1+\tau_i^L)w} \cdot \frac{P_i Y_i}{P_g Y_g} \\
\frac{K_i}{P_i Y_i} =& \frac{\epsilon_g - 1}{\epsilon_g} \cdot \frac{\alpha_g^K}{(1+\tau_i^K)R} \\
\frac{L_i}{P_i Y_i} =& \frac{\epsilon_g - 1}{\epsilon_g} \cdot \frac{\alpha_g^L}{(1+\tau_i^L)w}
\end{aligned}
$$

Plug these into TFPR_i and TFPR_g:

$$\begin{aligned}
\text{TFPR}_i &= \left(\frac{\epsilon_g - 1}{\epsilon_g} \cdot \frac{\alpha_s^K}{(1+\tau_i^K)R}\right)^{-\alpha_s^K} \left(\frac{\epsilon_g - 1}{\epsilon_g} \cdot \frac{\alpha_s^L}{(1+\tau_i^L)w}\right)^{-\alpha_s^L} \cdot (P_i Y_i)^{1-\alpha_s^K-\alpha_s^L} \\
&= \underbrace{(1+\tau_i^K)^{\alpha_s^K}(1+\tau_i^L)^{\alpha_L}\left(\frac{R}{\alpha_s^K}\right)^{\alpha_s^K}\left(\frac{w}{\alpha_s^L}\right)^{\alpha_s^L}\left(\frac{\epsilon_g}{\epsilon_g-1}\right)^{\alpha_s^K+\alpha_s^L}}_{\text{Same as CRS}}(P_i Y_i)^{1-\alpha_s^K-\alpha_s^L}
\end{aligned}$$

$$\begin{aligned}
\text{TFPR}_g &= \left(\sum_{i\in\mathcal{G}(g)} \frac{\epsilon_g - 1}{\epsilon_g} \cdot \frac{\alpha_s^K}{(1+\tau_i^K)R} \cdot \frac{P_i Y_i}{P_g Y_g}\right)^{-\alpha_s^K} \left(\sum_{i\in\mathcal{G}(g)} \frac{\epsilon_g - 1}{\epsilon_g} \cdot \frac{\alpha_s^L}{(1+\tau_i^L)w} \cdot \frac{P_i Y_i}{P_g Y_g}\right)^{-\alpha_s^L} \cdot (P_g Y_g)^{1-\alpha_s^K-\alpha_s^L} \\
&= \underbrace{\left(\sum_{i\in\mathcal{G}(g)} \frac{1}{1+\tau_i^K} \cdot \frac{P_i Y_i}{P_g Y_g}\right)^{-\alpha_s^K} \left(\sum_{i\in\mathcal{G}(g)} \frac{1}{1+\tau_i^L} \cdot \frac{P_i Y_i}{P_g Y_g}\right)^{-\alpha_s^L} \left(\frac{R}{\alpha_s^K}\right)^{\alpha_s^K}\left(\frac{w}{\alpha_s^L}\right)^{\alpha_s^L}\left(\frac{\epsilon_g}{\epsilon_g-1}\right)^{\alpha_s^K+\alpha_s^L}}_{\text{Same as CRS}} \\
&\quad \cdot (P_g Y_g)^{1-\alpha_s^K-\alpha_s^L}
\end{aligned}$$

In the code, we use an equivalent but easier formula because K_g and wL_g are observed. Follow HK, we define:

$$\begin{aligned}
\text{MPK}_g &\equiv \sum_{i\in\mathcal{G}(g)} \frac{P_i Y_i}{P_g Y_g (1+\tau_i^K)} = \frac{\epsilon_g}{\epsilon_g - 1} \cdot \frac{R}{\alpha_s^K} \cdot \frac{K_g}{P_g Y_g} \\
\text{MPL}_g &\equiv \sum_{i\in\mathcal{G}(g)} \frac{P_i Y_i}{P_g Y_g (1+\tau_i^L)} = \frac{\epsilon_g}{\epsilon_g - 1} \cdot \frac{w}{\alpha_s^L} \cdot \frac{L_g}{P_g Y_g}
\end{aligned}$$

Then we can write:

$$\frac{\text{TFPR}_i}{\text{TFPR}_g} = \underbrace{(1+\tau_i^K)^{\alpha_s^K}(1+\tau_i^L)^{\alpha_s^L}\text{MPK}_g^{\alpha_s^K}\text{MPL}_g^{\alpha_s^L}}_{\text{Same as CRS}}\left(\frac{P_i Y_i}{P_g Y_g}\right)^{1-\alpha_s^K-\alpha_s^L}$$

Nest, we derive the equilibrium market shares $\frac{P_i Y_i}{P_g Y_g}$. Using the optimal pricing rule, we can write the price ratio of two firms from the same nest as:

$$\frac{P_i}{P_j} = \left(\frac{A_j}{A_i}\right)^{\frac{1}{\alpha_s^L+\alpha_s^K}} \left(\frac{Y_i}{Y_j}\right)^{\frac{1}{\alpha_s^L+\alpha_s^K}-1} \left(\frac{1+\tau_i^K}{1+\tau_j^K}\right)^{\frac{\alpha^K}{\alpha_s^L+\alpha_s^K}} \left(\frac{1+\tau_i^L}{1+\tau_j^L}\right)^{\frac{\alpha^L}{\alpha_s^L+\alpha_s^K}}$$

Using demand side equation, $\frac{Y_i}{Y_j} = \left(\frac{P_i}{P_j}\right)^{-\epsilon_g}$, this can be rewritten as

$$\left(\frac{P_i}{P_j}\right)^{1+\epsilon_g\left(\frac{1}{\alpha_s^L+\alpha_s^K}-1\right)} = \left(\frac{A_j}{A_i}\right)^{\frac{1}{\alpha_s^L+\alpha_s^K}} \left(\frac{1+\tau_i^K}{1+\tau_j^K}\right)^{\frac{\alpha^K}{\alpha_s^L+\alpha_s^K}} \left(\frac{1+\tau_i^L}{1+\tau_j^L}\right)^{\frac{\alpha^L}{\alpha_s^L+\alpha_s^K}}$$

Demand side tells us, $\frac{P_iY_i}{P_jY_j} = \left(\frac{P_i}{P_j}\right)^{1-\epsilon}$, therefore

$$\frac{P_iY_i}{P_jY_j} = \left(\frac{A_j}{A_i}\right)^{\frac{1-\epsilon_g}{(1-\epsilon_g)(\alpha_s^L+\alpha_s^K)+\epsilon_g}} \left(\frac{1+\tau_i^K}{1+\tau_j^K}\right)^{\frac{\alpha^K(1-\epsilon_g)}{(1-\epsilon_g)(\alpha_s^L+\alpha_s^K)+\epsilon_g}} \left(\frac{1+\tau_i^L}{1+\tau_j^L}\right)^{\frac{\alpha^L(1-\epsilon_g)}{(1-\epsilon_g)(\alpha_s^L+\alpha_s^K)+\epsilon_g}}$$

Thus,

$$P_iY_i \propto \left(\frac{1}{A_i}\right)^{\frac{1-\epsilon_g}{(1-\epsilon_g)(\alpha_s^L+\alpha_s^K)+\epsilon_g}} (1+\tau_i^K)^{\frac{\alpha^K(1-\epsilon_g)}{(1-\epsilon_g)(\alpha_s^L+\alpha_s^K)+\epsilon_g}} (1+\tau_i^L)^{\frac{\alpha^L(1-\epsilon_g)}{(1-\epsilon_g)(\alpha_s^L+\alpha_s^K)+\epsilon_g}} \equiv W_i$$

Hence,

$$\frac{P_iY_i}{P_gY_g} = \frac{W_i}{\sum_{j\in g} W_j}$$

Plug the formula of $\frac{P_iY_i}{P_gY_g}$ in $\frac{\mathrm{TFPR}_i}{\mathrm{TFPR}_g}$:

$$\begin{aligned}\frac{\mathrm{TFPR}_i}{\mathrm{TFPR}_g} =& \Gamma_i \cdot \left(\frac{\Gamma_i}{A_i}\right)^{\theta_g(1-\alpha_g^K-\alpha_g^L)} \cdot \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{\Gamma_i}{A_i}\right)^{\theta_g}\right)^{-1} \\ &\cdot \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{\Gamma_i}{A_i}\right)^{\theta_g} \frac{1}{1+\tau_i^K}\right)^{\alpha_g^K} \cdot \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{\Gamma_i}{A_i}\right)^{\theta_g} \frac{1}{1+\tau_i^L}\right)^{\alpha_g^L} \end{aligned} \tag{7}$$

where

$$\begin{aligned}\Gamma_i \equiv& (1+\tau_i^K)^{\alpha_g^K}(1+\tau_i^L)^{\alpha_g^L} \\ \theta_g \equiv& \frac{1-\epsilon_g}{(1-\epsilon_g)(\alpha_g^K+\alpha_g^L)+\epsilon_g}\end{aligned}$$

Using Γ_i and θ_g, we can rewrite the equilibrium market shares as:

$$\frac{P_iY_i}{P_gY_g} = \frac{\left(\frac{\Gamma_i}{A_i}\right)^{\theta_g}}{\sum_{i\in\mathcal{G}(g)}\left(\frac{\Gamma_i}{A_i}\right)^{\theta_g}} \tag{8}$$

Combine Equation (6) and Equation (7) gives:

$$\text{TFP}_g = \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{\Gamma_i}{A_i}\right)^{\theta_g} \right)^{\frac{\epsilon_g}{\epsilon_g - 1}} \cdot \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{\Gamma_i}{A_i}\right)^{\theta_g} \frac{1}{1+\tau_i^K} \right)^{-\alpha_g^K} \cdot \left(\sum_{i\in\mathcal{G}(g)} \left(\frac{\Gamma_i}{A_i}\right)^{\theta_g} \frac{1}{1+\tau_i^L} \right)^{-\alpha_g^L} \tag{9}$$

When there are no input distortions, then τ_i^K and τ_i^L are set to 0. From Equation (8):

$$\frac{P_i^* Y_i^*}{P_g^* Y_g^*} = \frac{A_i^{-\theta_g}}{\sum_{i\in g} A_i^{-\theta_g}}$$

From Equation (9):

$$\text{TFP}_g^* = \left(\sum_{i\in\mathcal{G}(g)} A_i^{-\theta_g} \right)^{-\frac{1}{\theta_g}}$$

B HK's data, our data, and Chinese Yearbooks

Both HK and we use the annual survey data of Chinese industries. Ours is a newer version acquired via Peking University. Table 7 and Table 8 show how much the aggregates of the two ASM data deviate from the counterpart macro variables published in China Statistical Yearbooks (CSYs) reported as percentage shares of those variables in CSYs. HK only have 1998-2005 so Table 8 only reports these years. The differences between our data and CSYs are mostly around or below 2% while those between HK's data and CSYs are around $10 - 20\%$. Our data contains around $0.05 - 0.1\%$ more firms than CSYs in each year except for 2004 and 2008 while HK's data contains around 20% less firms in 1998-2002 and around 10% less in 2003-2005.

Table 7: Our data statistics in comparison with China Statistical Yearbook: ratio (%)

Year	Number of firms	Sales	Output	Value added	Employment	Net value of fixed assets	Export	profits
1998	0.05	0.41	0.38	0.41	-8.56	1.48	0.58	-2.76
1999	0.04	0.94	1.02	0.92	0.46	-2.21	1.19	0.20
2000	0.06	0.54	0.51	0.45	0.39	-1.31	0.11	0.09
2001	0.06	0.89	1.26	1.14	0.54	-1.50	0.81	1.91
2002	0.07	0.84	0.83	0.83	0.37	-1.57	0.16	0.64
2003	0.10	1.80	1.78	1.88	1.00	-1.41	1.59	2.32
2004	-0.54	0.78	0.74	5.20	0.98	-2.72	1.06	1.95
2005	0.09	1.24	1.22	1.30	1.14	-2.76	1.17	1.39
2006	0.12	1.38	1.18	1.12	0.64	-3.01	3.05	1.23
2007	0.13	1.95	1.64	2.14	1.52	-3.06	1.96	3.48
2008	-3.30	-0.74	-1.40		-2.74	-6.02	-0.46	-1.28

Notes: all the variables are from from the latest available yearbook issue.
Export data of China Statistical Yearbook is from Brandt et al. (2014).

Table 8: HK's data statistics in comparison with China Statistical Yearbook: ratio (%)

Year	Number of firms	Sales	Value added	Employment	Net value of fixed assets	Export
1998	-27.87	-14.74	-20.19	-23.24	-16.69	-19.12
1999	-26.09	-12.22	-18.62	-19.84	-15.23	-14.09
2000	-23.14	-9.04	-12.81	-20.92	-2.53	-10.47
2001	-22.17	-10.44	-13.85	-19.17	-2.57	-11.41
2002	-19.00	-8.19	-11.54	-14.64	0.15	-9.13
2003	-14.15	-5.95	-5.96	-10.24	2.14	-6.34
2004	-9.44	-4.95	-10.34	33.13	-2.94	
2005	-8.40	-4.63	-12.66	-6.19	-2.74	-4.75

Notes: all the variables are from from the latest available yearbook issue.
Export data of China Statistical Yearbook is from Brandt et al. (2014).