The Aggregate Labor Share and Distortions in China *

Xiaoyue Zhang† Junjie Xia ‡

March 24, 2025

Abstract

This paper quantifies the impact of distortions that hinder firms from using profit-maximizing amount of capital and labor on the labor share in the aggregate revenue of Chinese manufacturing, mining, and public utilities. This quantification takes into account heterogeneous productivity, technology, and demand elasticities across firms. Using parameters estimated from firm-level data, we find that removing the distortions and holding the aggregate labor and capital supply fixed would raise the labor share by 24 percentage points. This increase in the labor share is driven by a 57% increase in the wage.

Keywords: Distortions, aggregate labor share, latent market structure, firm heterogeneity

*We are grateful to Jaap H. Abbring, Christoph B. T. 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 Zhang is also thankful for the never-ending support and confidence from Edi Karni.

†DIW Berlin, MIT Sloan, corresponding author, email: xzhang@diw.de and xyzhang1@mit.edu

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

1 Introduction

This paper quantifies how the labor share in the aggregate revenue of Chinese manufacturing, mining, and public utilities changes when firms fail to use their profit-maximizing amount of capital and labor. Distortions such as a lack of access to financial credits, adjustment frictions, and regulations can constrain the use. Restuccia and Rogerson (2008), Hsieh and Klenow (2009), and Zhang and Xia (2022) document that these distortions are pervasive, and can generate 30% to 50% decreases in total factor productivity in large economies such as China. Removing these distortions would reallocate labor and capital to more productive use. However, recent evidence suggests that allocating production factors to more productive firms reduce the aggregate labor share of the U.S. firms (Autor et al. (2020)). Whether the aggregate labor share in China would also be reduced after removing the distortions is unclear. Our estimation shows that holding the aggregate labor and capital fixed, the labor share would increase by 24 percentage points and that this increase is driven by a 57% increase in the wage.

We divide the distortions into input and output distortions. The input distortions distort the observed labor-capital ratios from the theoretical profit-maximizing ratios. They are the firm-level wedges between the observed cost ratios and the production-elasticity ratios of labor and capital. The output distortions affect firm sizes by proportionally changing the use of capital and labor without altering the capital-labor ratios. Markups predicted by demand elasticities are excluded from the distortions.

While removing the input and output distortions generates efficiency gains, it does not guarantee a higher aggregate labor share. Changes in the aggregate labor share depends on whether distortions cause firms to substitute away from labor and how resources are allocated across firms in the general equilibrium. In the special case where the distortions impede firms' capital use but do not restrain labor use, removing the distortions would cause firms to substitute away from labor, lower the aggregate labor demand, push down the wage, and consequently lower the aggregate labor share. Similarly, the aggregate labor share would increase when the distortions are in the opposite direction.

Our paper provides estimates of firm-level input and output distortions and quantifies the impact of these distortions on the aggregate labor share. Due to the firm-level estimates, we are able to investigate how the distortions affect the aggregate labor share, which type of the distortions play the main role, how the labor allocation changes, and how market concentration is affected.

We model firm heterogeneity using firm-specific productivity and distortions and industry-specific technology. Demand elasticities are allowed to vary within and across industries. On the supply side, we use Cobb-Douglas production functions with constant returns to scale on labor and capital.[1] Production functions are the same within industry except for firm-specific productivity but differ across industries. On the demand side, we assume nested constant elasticities of substitutions (nested CES). The nest structure is latent up to industry categories. Firms inside the same nest face the same constant elasticity of substitution, but the elasticities vary across nests. The elasticities of substitution across nests equal 1. We choose a nested CES demand with a latent nest structure to keep the demand side flexible yet parsimonious. Cobb-Douglas production functions offer us simple general equilibrium conditions for estimating firm heterogeneity. We assume a perfectly competitive factor market where all the firms pay the same wage and capital rental rate.

We use the Chinese Annual Survey of Industry in 2005 to estimate the demand and supply parameters. The survey covers all the state-owned firms and above-scale non-state firms, i.e. non-state firms with revenues above 5 million RMB ($600,000). We do not observe the nest structure in the nested CES demand, but we observe firms' industry categories. To identify the latent nest structure, we assume that a nest belongs to only one industry and that the logarithm of firms' revenue-cost ratios are firms' markups predicted by their demand elasticities plus an idiosyncratic error term, which follows a normal distribution within a nest. Under these assumptions, the distribution of firm-level revenue-cost ratios within an industry is a mixture normal distribution with an unknown number of components. We estimate firms' demand elasticities and the latent nest structure by estimating a mixture normal distribution for each industry. We then use the observed firm-level expenditure ratios of labor and capital to estimate production elasticities of substitution assuming that the mode of input distortions within an industry is 0 and that the production elasticities are the same within an industry. We choose the mode because we are concerned that the industry-level distributions are asymmetric and may have thick tails. Finally, we use the estimated demand and supply parameters to calculate the input and output distortions using the firm-level labor and capital expenditure shares out of value added.

Our results show that the aggregate labor share would increase by 24 percentage points when firms can utilize the profit-maximizing amount of capital and labor. Under the assumption of fixed aggregate capital and labor, the increase in the labor share is driven by a 57% increase in the wage. This increase is almost entirely driven by the output distortions. At the firm level, about 74% of the firms are too small due to the output distortions. In terms of the distortions on labor-capital ratios, about half of the firms overuse labor and

[1] Due to the unitary elasticity of substitution of Cobb-Douglas production functions, the changes in the aggregate labor share is driven by removing the gaps between the firm-level observed labor shares and the theoretical profit-maximizing labor shares and by the reallocation of capital and labor across firms.

the other half underuse labor relative their capital use. The joint impact of the input and output distortions at the firm level is that 76% of the firms' labor shares are lower than their profit-maximizing values. Intuitively, when removing the distortions, the 76% firms' demand for labor increases, which aggregates to a higher aggregate labor demand. Equilibrium wage increases to clear the market and labor is reallocated to firms' with a higher increase in labor demand. Because the aggregate demand for labor increases disproportionally more than the increase in the aggregate production, the aggregate labor share increases. In terms of the reallocation across firms, most firms shrink but 1% of the firms grow 8 times larger or more. As a consequence, market concentration increases. The average industry Herfindahl-Hirschman Index after removing the distortions is 5.5 times as large as the one with the distortions.

Scholars have long been interested in whether there is a trade-off between reallocating production resources to improve efficiency versus to increase the aggregate labor share. Autor et al. (2020) find that when more productive firms grow, the aggregate labor share declines because more productive firms tend to have lower labor shares. Similar concerns exist in the discussion on whether the aggregate labor share will decline when new technology can replace labor (Acemoglu (2003), Acemoglu and Restrepo (2018), and Jones and Liu (2024)). Our paper shows that both the efficiency and the aggregate labor share would increase if Chinese firms can utilize profit-maximizing amount of capital and labor.

The classical misallocation model in Hsieh and Klenow (2009) relies on a rigid nested CES demand structure in which a nest is an industry and that all the nests have the same demand elasticity. Recent studies such as Haltiwanger et al. (2018) point out that the underlying assumptions of the rigid nested CES demand are likely violated and that the quantified efficiency costs of misallocation are biased. Despite attempts to generalize the demand assumptions (Liang (2023) compares the predicted efficiency costs using alternative demand structures), it is difficult to balance a demand structure's flexibility and its ability to provide general equilibrium conditions simple enough for estimation using firm-level data. We follow Baqaee and Farhi (2020)'s generalization by allowing arbitrary number of nests in a nested CES demand. We take Baqaee and Farhi (2020) one step further by estimating the nest structure and nest-specific demand elasticities using firm-level data.

The remainder of the paper is organized as follows. We introduce the data set in Section 2 and describe our theoretical model in Section 3. Section 4 explains our estimation strategy and Section 5 presents our results. We discuss the robustness of our results in Section 6 . Section 7 concludes. The Appendix provides the details of data cleaning, the derivation of theoretical results of Section 3, and a discussion of alternative ways of inferring firm-level markups.

2 Data

Our data source is the annual survey conducted by the National Bureau of Statistics (NBS) in China in 2005.[2] The annual surveys have been used by previous studies including Hsieh and Klenow (2009), Brandt et al. (2012), and David and Venkateswaran (2019). The surveys include all the state-owned firms and all the non-state firms with revenues above 5 million RMB ($600,000). The survey in 2005 contains about 230,000 firms. It covers manufacturing, mining, and public utilities. The industry classification used in this paper has 523 industries.

The data contains information on firm-level value added, wage expenditure, fixed assets, revenues, and costs. When cleaning the data, we drop unreasonable observations in terms of accounting as in Brandt et al. (2012), such as negative value added, negative debts, and negative revenues. Appendix A lists all the criteria that trigger dropping an observation. We calculate the net present value of depreciated real capital also following Brandt et al. (2012). The net present value of depreciated real capital is our measure of capital. Value added is the monetary value of a firm's products produced in the year after netting out intermediate input costs. The monetary value per product is the average sales price of the product in the year. Revenues in the data are the monetary value of the products sold by a firm in the year, which can differ from the value of products produced in the year. We trim the 1% tails of value added, labor and capital shares of value added, revenue-cost ratios, capital, and labor expenditure.

Table 1: Summary statistics weighted by firm-level value added

	Mean	Standard Deviation	10th Percentile	1st Quartile	Median	3rd Quartile	90th Percentile
labor share	0.26	0.28	0.04	0.08	0.17	0.34	0.55
adj. labor share	0.42	0.26	0.11	0.21	0.38	0.60	0.78
capital share	0.17	0.23	0.02	0.04	0.10	0.21	0.38
cost share	0.59	0.37	0.18	0.32	0.55	0.79	1.02
cost/revenue	0.83	0.13	0.67	0.78	0.87	0.92	0.96

Notes: Total number of firms is 229,282. Capital rental rate is assumed 0.2. Cost share is adjusted labor share plus capital share.

Table 1 provides the summary statistics of firm-level labor and capital shares out of value added, cost shares of labor and capital out of value added, and cost-revenue ratios. The summary statistics are weighted by value added. The unweighted summary statistics are in Table 5 in Appendix A. The labor share is the firm-level labor expenditure divided by value added. The annual survey collects firms' wage expenditure, but it does not collect firms'

[2]We acquired the data through Peking University's data center.

expenditure on non-wage labor compensation. The aggregate labor share calculated using the reported wage expenditure in the data is 26%, but the Bai and Qian (2010) estimate that the aggregate labor share of manufacturing, mining, and public utilities in China in 2005 is 42%. We assume a constant ratio between unobserved non-wage compensation to wage compensation at the firm level and adjust the firm-level labor expenditure so that the aggregate labor share out of value added is 42%. This assumption of a constant ratio is used in Hsieh and Klenow (2009) to solve the same problem.

The capital share in Table 1 is firm-level capital multiplied by the capital rental rate and divided by value added. We assume the capital rental rate to be 0.2 following Bai Chong-En et al. (2006) and Sun et al. (2011). The aggregate capital share in our data is 17%. Chinese Year Book for 2005 reports that the naional income share of capital is 15%, but the national income includes construction sector, service sector, and the agriculture sector in addition to manufacturing, mining, and public utilities sector in our data. The cost share in Table 1 is the sum of capital and adjusted labor shares.

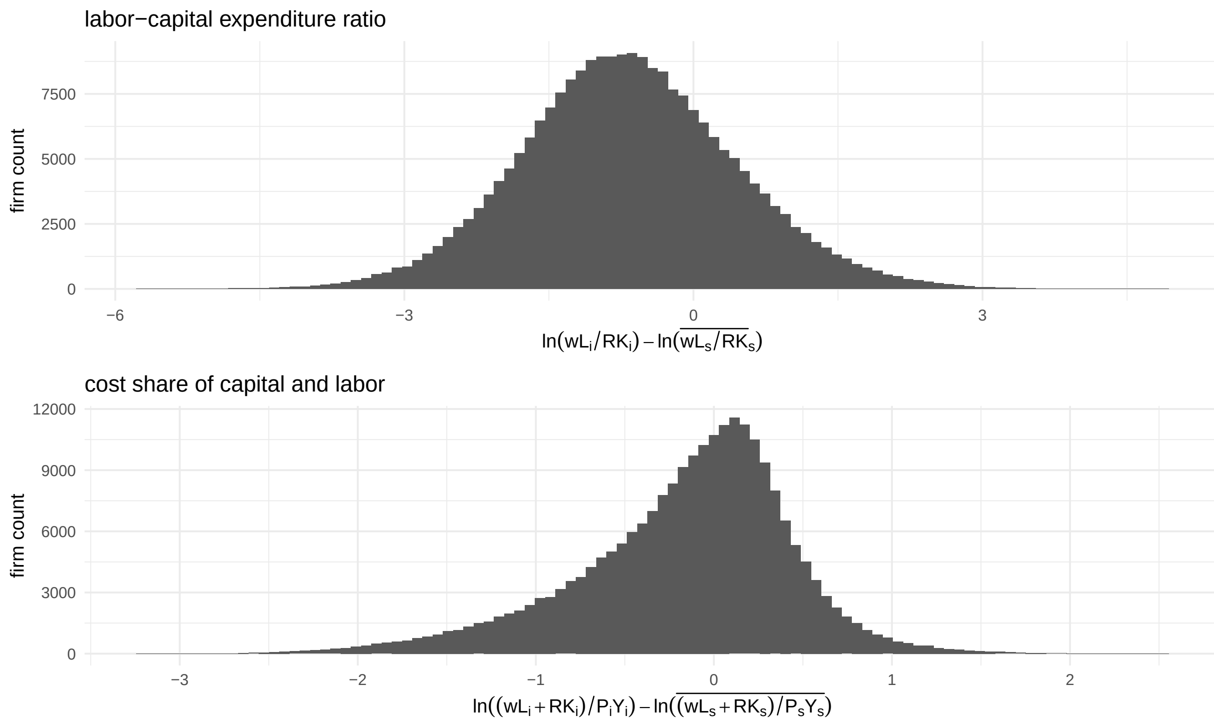
The last row in Table 1 is firms' cost-revenue ratios calculated using firms' reported total costs and total revenues. They are different from the cost shares, which is the share of firms' capital-and-labor expenditure out of their value added. The last two rows of Table 1 show that firms' labor-and-capital cost shares are on average smaller than cost-revenue ratios and that the variation in firms' labor-and-capital cost shares is larger than the one in the cost-revenue ratios.

Figure 1 demonstrates the variation in firm-level labor-capital expenditure ratios and firm-level cost shares of labor and capital out of value added. For comparison across industries, the expenditure ratios and cost shares are normalized by industry averages. The distribution of firms' normalized labor-capital expenditure ratios looks symmetric, but the mode is negative. The distribution of firms' normalized labor-and-capital cost shares is skewed to the left with a mode slightly above 0. The dispersion of the normalized cost shares is smaller than the one of the normalized expenditure ratios.

Figure 2 continues investigating the the within-industry dispersion in the expenditure ratios and the cost shares. It plots the distribution of within-industry coefficients of variation for labor-capital expenditure ratios and labor-and-capital cost shares. The within-industry coefficients of variation measure the size of the variation within an industry. Figure 2 shows that the variation differs across industries. Similar to Figure 1, the variation is larger for the expenditure ratios than for the cost shares.

Although the distribution of the normalized labor-capital expenditure ratios in Figure 1 looks symmetric, the symmetry is due to pooling all the industries together. Figure 3 displays the industry-level skewness and kurtosis of the logarithm of firms' labor-capital expenditure

FIGURE 1: Distribution of the normalized labor-capital expenditure ratios and labor-and-capital cost shares



in industry s.

Notes: $\overline{wL_s/RKs}$ is the average of firm-level labor-capital expenditure ratios wL_i/RK_i in industry s. $\overline{(wL_s + RK_s)/P_sY_s}$ is the average of firm-level cost shares of labor and capital $\overline{(wL_i + RK_i)/P_iY_i}$

Figure 2: Within-industry coefficients of variation of the expenditure ratios and the cost shares

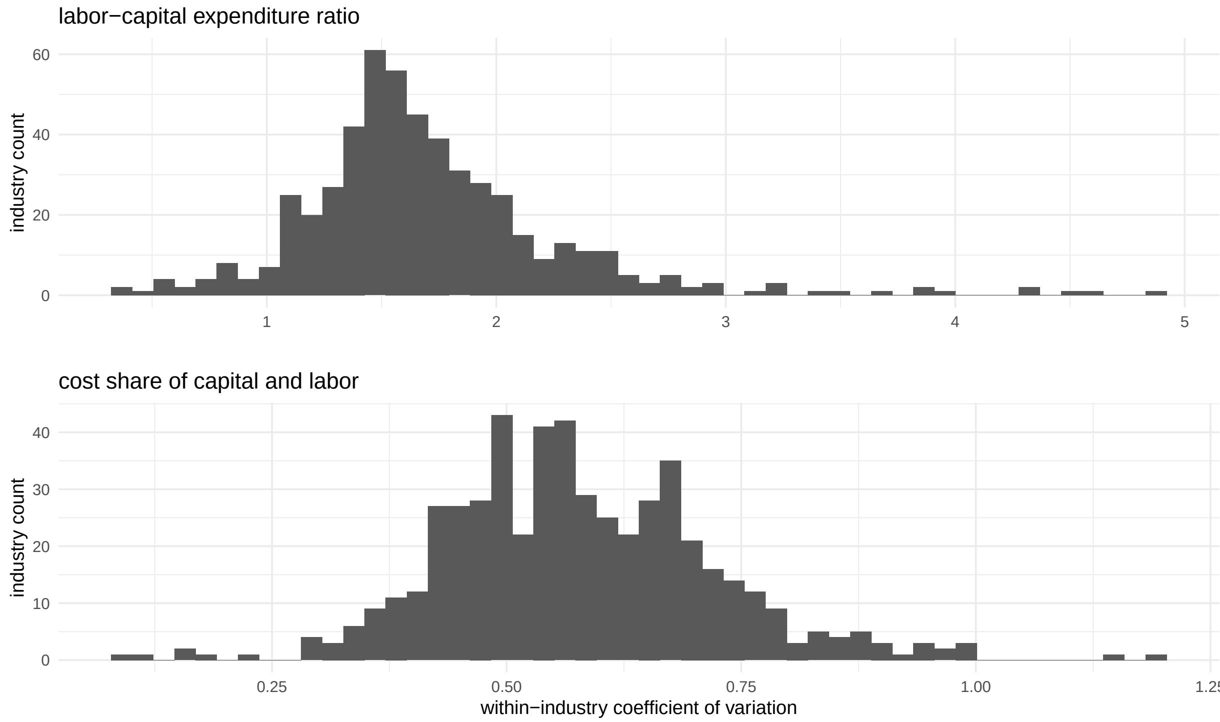


Figure 3: Within-industry skewness and kurtosis of the capital-labor expenditure ratios

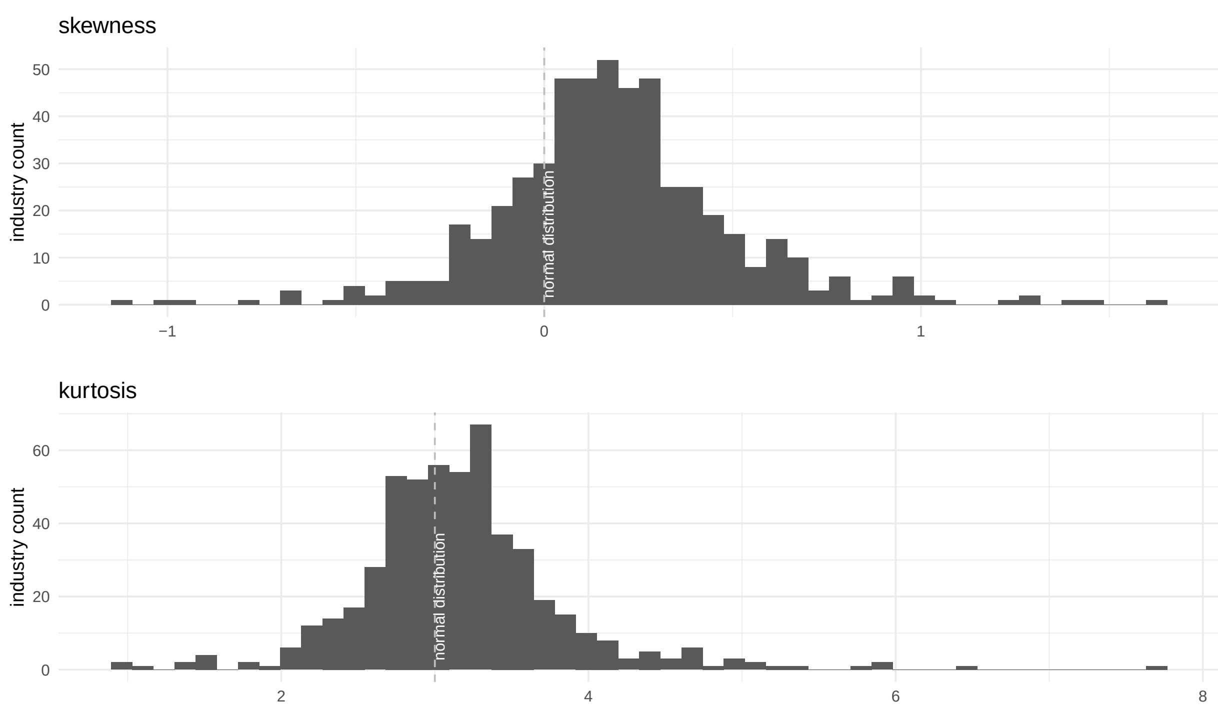
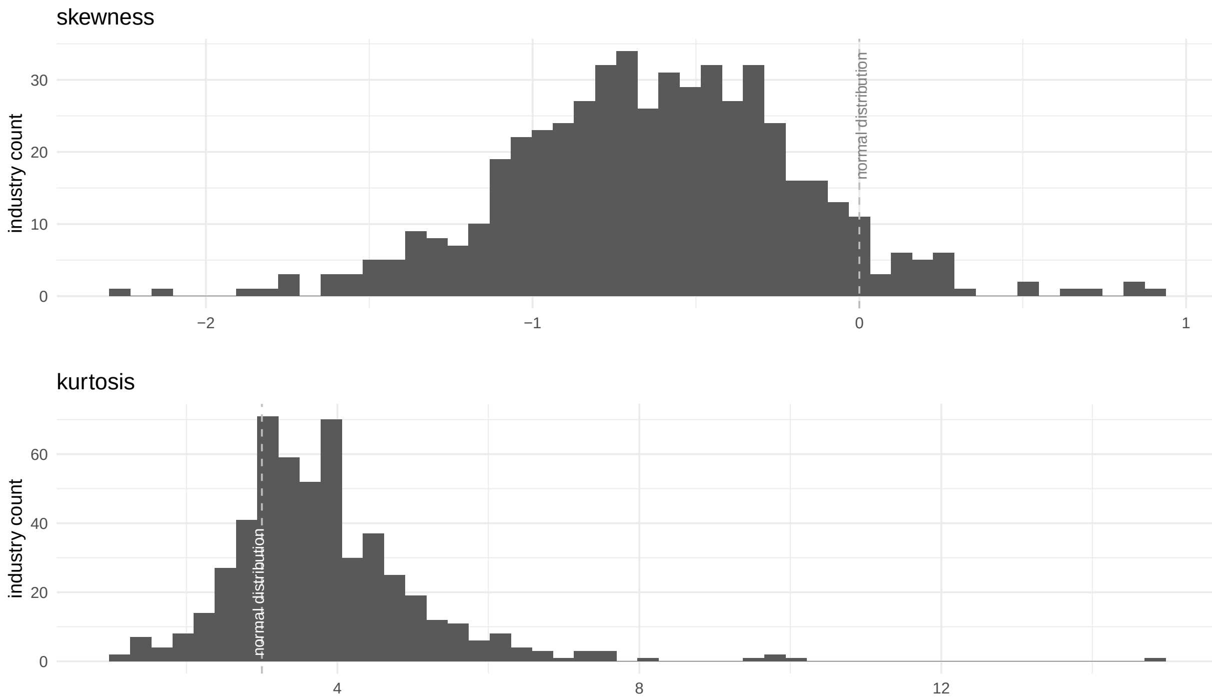


Figure 4: Within-industry skewness and kurtosis of the capital-labor cost shares



ratios. 76% of the industries are skewed to the right and 60% have tails thicker than normal distributions. Figure 4 is the same plot as Figure 3 but for the logarithm of firms' labor-capital cost shares. 94% of the industries are skewed to the left and 76% have tails thicker than normal distributions.

3 Model

In this section, we first introduce the model which we use to describe the economy. The formal definition of the input and output distortions are provided in this part. We then derive the general equilibrium conditions and provide the predicted changes in the aggregate labor share. In the last part of the section, we offer two special cases to demonstrate how the distortions affect the economy.

3.1 Model setup

The economy has S number of industries $\{s_1, \cdots, s_S\}$. Firms inside an industry have the same Cobb-Douglas production function upto firm-specific productivity A_i:

$$F_i(K, L) = A_i K^{1-\alpha_{s(i)}} L^{\alpha_{s(i)}} \tag{1}$$

where i denotes firm i and $s(i)$ is the industry firm i belongs to.

Without distortions, firm i's profit is its revenue net of its expenditure on capital and labor:

$$\Pi_i(P, K, L) = F_i(K, L)P - wL - RK \tag{2}$$

where w is the wage and R is the capital rental rate. Firm i takes wage and capital rental rate as given and chooses price P, capital K, and labor L to maximize its profits $\Pi_i(P, K, L)$. The wage and the capital rental rate are the same for all the firms in the economy.

On the demand side, there is a representative consumer with a nested constant elasticities of substitutions (nested CES) utility:

$$\bar{Y} \equiv \prod_g \bar{Y}_g^{\beta_g} \tag{3}$$

where $\bar{Y}_g = \left(\sum_{i \in \mathcal{G}(g)} Y_i^{\frac{\epsilon_g - 1}{\epsilon_g}} \right)^{\frac{\epsilon_g}{\epsilon_g - 1}}$. Y_i denotes the amount of firm i's product consumed by the representative consumer. $\mathcal{G}(g)$ is the set of firms in nest g, and they face the same demand elasticity ϵ_g.

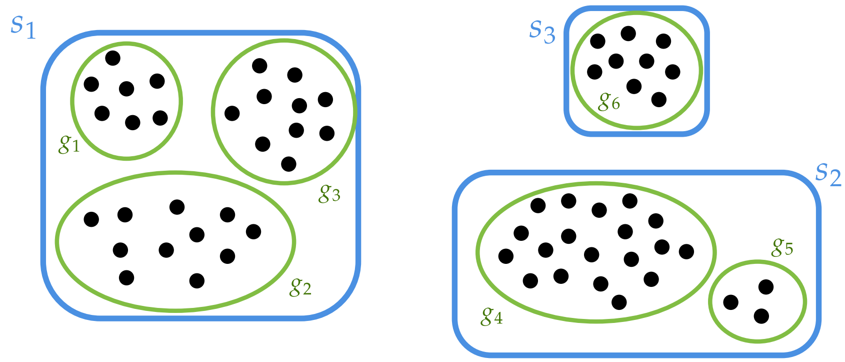
A nest can only belong to one industry but an industry can have multiple nests:

$$\bigcup_{g \in \{g_1, \cdots, g_s\}} \{i | g(i) = g\} = \mathcal{S}(s) \tag{4}$$

$$\{i | g(i) = g_l\} \bigcap \{i | g(i) = g_m\} = \emptyset \text{ , for } \forall g_l \neq g_m \tag{5}$$

where $\{g_1, \cdots, g_s\}$ is the set of nests which contain firms in industry s. $g(i)$ denotes the nest of firm i, and $\mathcal{S}(s)$ is the set of firms in industry s. Equation (4) means that the set of firms in industry s is the union of firms in nests $\{g_1, \cdots, g_s\}$. Equation (5) means that there is no overlap between any two nests. Figure 5 provides an illustrative example of a three-industry economy. Industry s_1 has 3 nests, g_1, g_2, and g_3. Industry s_2 has 2 nests g_4 and g_5. Industry s_3 has 1 nest g_6.

FIGURE 5: An illustrative example of a three-industry economy



Notes: Black dots are firms, green circles are nests, and blue squares are industries.

The total amount of labor and capital in the economy is fixed:

$$\sum_i L_i = \bar{L} \ , \ \sum_i K_i = \bar{K} \tag{6}$$

where $\sum\limits_i$ sums over all the firms in the economy.

In the general equilibrium, on the supply side, each firm takes wage w and capital rental rate R as given and maximizes their profits by choosing the amount of capital and labor to use and setting their prices under Chamberlinian monopolistic competition. The Chamberlinian monopolistic competition means that firms do not take into account the impact of their pricing and production decisions on other firms' pricing and production decisions. On the demand side, the representative consumer takes each product's price P_i as given and maximizes its utility $\bar{Y}$ net of the expenditure on the products, $\sum_i P_i Y_i$. Wage w and capital rental rate R clear the market for capital and labor. Formally, the economy's general equilibrium is representative by a vector $\{\{P_i, K_i, L_i\}_i, w, R\}$ that satisfies

$$\{Y_i\}_i = \arg\max_{\{y_i\}_i} \prod_g \left[\left(\sum_{i \in \mathcal{G}(g)} y_i^{\frac{\epsilon_g - 1}{\epsilon_g}} \right)^{\frac{\epsilon_g}{\epsilon_g - 1}} \right]^{\beta_g} - \sum_i P_i y_i \tag{7}$$

$$\{P_i, K_i, L_i\} = \arg\max_{P,K,L} PF_i(K, L) - wL - RK \text{ for } \forall i \tag{8}$$

$$y_i = F_i(K_i, L_i) \tag{9}$$

$$\bar{L} = \sum_i L_i \tag{10}$$

$$\bar{K} = \sum_i K_i \tag{11}$$

We use capital letters without subscripts to denotes variables in functions and capital letters with subscripts to denotes the values these variables take. In the representative consumer's problem, i.e. Equation (7), we use lower-case letters with a subscript, y_i, to denote the variable version of the amount of product i consumed and capital-letter Y_i to denote the value that y_i takes because we need to differentiate production by different firms. Equation (8) is the firms' problem. Equation (9), (10), and (11) clear the goods market, the labor market, and the capital market.

Without distortions, firm i's equilibrium expenditure ratio of capital and labor and its equilibrium input expenditure share of capital and labor are:

$$\frac{w^* L_i^*}{R^* K_i^*} = \frac{\alpha_{s(i)}}{1 - \alpha_{s(i)}}$$

$$\frac{w^* L_i^* + R^* K_i^*}{P_i^* Y_i^*} = \frac{\epsilon_{g(i)} - 1}{\epsilon_{g(i)}}$$

When there are distortions in the economy, the distortions $1 + \tau_i^I$ and $1 - \tau_i^Y$ create gaps between the left and right hand sides of the equations so that the first-order conditions of the firms' problem do not hold:

$$\frac{wL_i}{RK_i} = (1 + \tau_i^I) \cdot \frac{\alpha_{s(i)}}{1 - \alpha_{s(i)}}$$

$$\frac{wL_i + RK_i}{P_i Y_i} = (1 - \tau_i^Y) \frac{\epsilon_{g(i)} - 1}{\epsilon_{g(i)}}$$

Rearranging the equations gives the definition of the input and output distortions:

$$\ln(1+\tau_i^I) \equiv \ln\left(\frac{wL_i}{RK_i}\right) - \ln\left(\frac{\alpha_{s(i)}}{1-\alpha_{s(i)}}\right) \tag{12}$$

$$\ln(1-\tau_i^Y) \equiv \ln\left(\frac{wL_i+RK_i}{P_iY_i}\right) - \ln\left(\frac{\epsilon_{g(i)}-1}{\epsilon_{g(i)}}\right) \tag{13}$$

In other words, the input distortions are the firm-level wedges between firms' expenditure ratios and production elasticities ratios of labor and capital. The output distortions are the firm-level wedges between the labor-and-capital cost shares and the inverse of the markups predicted by the demand elasticities. We require $1+\tau_i^I > 0$ and $1-\tau_i^Y > 0$ because the expenditure ratios and the cost shares are positive. Firm i overuses labor relative to capital when $1+\tau_i^I > 1$ and underuses labor relative to capital when $0 < 1+\tau_i^I < 1$. Firm i is too small, i.e. proportionally underuses both labor and capital, when $0 < 1-\tau_i^Y < 1$ and too big when $1-\tau_i^Y > 1$.

The decomposition of firm i's revenue is:

$$\underbrace{P_iY_i}_{\text{revenue}} = \underbrace{wL_i+RK_i}_{\substack{\text{expenditure on}\\ \text{labor and capital}\\ \text{(contains input}\\ \text{distortions)}}} + \underbrace{P_iY_i\frac{1}{\epsilon_{g(i)}}(1-\tau_i^Y)}_{\text{operating surplus}} + \underbrace{\tau_i^Y P_iY_i}_{\substack{\text{collected in}\\ \text{output distortions}}} \tag{14}$$

We call $P_iY_i\frac{1}{\epsilon_{g(i)}}(1-\tau_i^Y)$ firm i's operating surplus to distinguish it from firm i's realized profit. The operating surplus is the profits earned due to demand elasticities $\epsilon_{g(i)}$. Profits are revenues net of costs. Whether profits equal operating surplus depends on whether revenues collected in output distortions are costs or profits.

Aggregating Equation (14) over all the firms in the economy gives the decomposition of the aggregate revenue in this economy :

$$\underbrace{\bar{P}\bar{Y}}_{\substack{\text{aggregate}\\ \text{revenue}}} = \underbrace{wL+RK}_{\substack{\text{aggregate expenditure}\\ \text{on labor and capital}\\ \text{(contains input}\\ \text{distortions)}}} + \underbrace{\sum_i P_iY_i\frac{1}{\epsilon_{g(i)}}(1-\tau_i^Y)}_{\text{aggregate operating surplus}} + \underbrace{\sum_i \tau_i^Y P_iY_i}_{\substack{\text{collected in}\\ \text{output distortions}}} \tag{15}$$

3.2 Predicted changes in the aggregate labor share

The predicted change in the aggregate labor share due to removing the distortions is the difference in the aggregate labor share with and without the distortions. The aggregate labor share with the distortions is:

$$
\begin{aligned}
\frac{w\bar{L}}{\bar{P}\bar{Y}} &= \sum_g \frac{\bar{P}_g \bar{Y}_g}{\bar{P}\bar{Y}} \cdot \frac{w\bar{L}_g}{\bar{P}_g \bar{Y}_g} \\
&= \sum_g \frac{\bar{P}_g \bar{Y}_g}{\bar{P}\bar{Y}} \sum_{i \in \mathcal{G}(g)} \frac{wL_i}{P_i Y_i} \cdot \frac{P_i Y_i}{\bar{P}_{g(i)} \bar{Y}_{g(i)}} \\
&= \sum_g \frac{\bar{P}_g \bar{Y}_g}{\bar{P}\bar{Y}} \sum_{i \in \mathcal{G}(g)} (1-\tau_i^Y) \cdot \frac{\alpha_{s(i)} \cdot (1+\tau_i^I)}{1+\alpha_{s(i)}\tau_i^I} \cdot \frac{\epsilon_{g(i)} - 1}{\epsilon_{g(i)}} \cdot \frac{P_i Y_i}{\bar{P}_{g(i)} \bar{Y}_{g(i)}}
\end{aligned}
$$

where $\bar{P}\bar{Y} \equiv \sum_i P_i Y_i$ is the aggregate revenue, $\bar{P}_g \equiv \sum_{i \in \mathcal{G}(g)} \frac{P_i Y_i}{\bar{Y}_{g(i)}}$ is the price index of nest g, $\bar{L}_g \equiv \sum_{i \in \mathcal{G}(g)} L_i$ is the total amount of labor used in nest g. The first two equations use the fact that $w\bar{L} = \sum_g w\bar{L}_g$ and $w\bar{L}_g = \sum_{i \in \mathcal{G}(g)} wL_i$. The third equation uses the definition of the distortions in Equation (12) and (13).

Solving the representative consumer's problem in Equation (7) gives nests' market shares and firms' shares within nests:

$$
\frac{\bar{P}_g \bar{Y}_g}{\bar{P}\bar{Y}} = \beta_g \tag{16}
$$

$$
\frac{P_i Y_i}{\bar{P}_{g(i)} \bar{Y}_{g(i)}} = \frac{P_i^{1-\epsilon_{g(i)}}}{\sum_{i \in \mathcal{G}(g(i))} P_i^{1-\epsilon_{g(i)}}} \tag{17}
$$

Solving the general equilibrium gives us the aggregate labor share under distortion:

$$
\frac{w\bar{L}}{\bar{P}\bar{Y}} = \sum_g \sum_{i \in \mathcal{G}(g)} \beta_g \cdot \underbrace{(1-\tau_i^Y) \frac{\epsilon_{g(i)} - 1}{\epsilon_{g(i)}} \cdot \frac{\alpha_{s(i)}(1+\tau_i^I)}{1+\alpha_{s(i)}\tau_i^I}}_{\text{firm-level labor shares}} \cdot \underbrace{\frac{\gamma_i^{1-\epsilon_{g(i)}}}{\sum_{i \in \mathcal{G}(g(i))} \gamma_i^{1-\epsilon_{g(i)}}}}_{\text{firm-level market shares}} \tag{18}
$$

where

$$
\gamma_i \equiv \underbrace{\frac{1}{1-\tau_i^Y}}_{\text{(a)}} \cdot \frac{1}{A_i} \underbrace{\left[(1+\tau_i^I)^{-\alpha_{s(i)}} \cdot (1-\alpha_{s(i)}) + (1+\tau_i^I)^{1-\alpha_{s(i)}} \cdot \alpha_{s(i)}\right]}_{\text{(b)}} \tag{19}
$$

Removing the distortions reallocates capital and labor across firms. The reallocation's impact on the aggregate labor share can be interpreted in two parts: how individual firms' labor shares change and how their market shares change. Equation (18) and Equation (19) show that the two parts are affected by both input and output distortions.

A closer examination of Equation (19) offers insights on how the distortions alter firms' market shares. In Equation (19), Part (a) is the impact of the output distortions, which serve as a factor on firms' productivity A_i since they distort capital and labor use proportionally. When $1-\tau_i^Y < 1$, the output distortion causes firm i to appear less productive than it actually is. Part (b) shows that the input distortions affect firm sizes by raising marginal costs.

The aggregate labor share without distortions is:

$$\frac{w^*\bar{L}}{\bar{P}^*\bar{Y}^*} = \sum_g \beta_g \frac{\epsilon_g - 1}{\epsilon_g} \alpha_{s(g)} \tag{20}$$

where $s(g)$ denotes the industry that nest g belongs to. The predicted change in the aggregate labor share due to removing the distortions is:

$$\begin{aligned}\frac{w^*\bar{L}}{\bar{P}^*\bar{Y}^*} - \frac{w\bar{L}}{\bar{P}\bar{Y}} =& \sum_g \beta_g \frac{\epsilon_g - 1}{\epsilon_g} \alpha_{s(g)} \\ &- \sum_g \sum_{i \in \mathcal{G}(g)} \beta_g \cdot (1-\tau_i^Y) \frac{\epsilon_{g(i)} - 1}{\epsilon_{g(i)}} \cdot \frac{\alpha_{s(i)}(1+\tau_i^I)}{1+\alpha_{s(i)}\tau_i^I} \cdot \frac{\gamma_i^{1-\epsilon_{g(i)}}}{\sum_{i \in \mathcal{G}(g(i))} \gamma_i^{1-\epsilon_{g(i)}}}\end{aligned} \tag{21}$$

Firm i's equilibrium marginal cost mc_i is:

$$\text{mc}_i = R \cdot \underbrace{A_i^{-1} \left[\frac{R}{w} \cdot \frac{\alpha_{s(i)}}{1-\alpha_{s(i)}} \cdot (1+\tau_i^I) \right]^{-\alpha_{s(i)}}}_{\text{(a)}} \cdot \underbrace{\left[1 + \frac{\alpha_{s(i)}}{1-\alpha_{s(i)}} (1+\tau_i^I) \right]}_{\text{(b)}} \tag{22}$$

Part (a) in Equation (22) is how much capital is needed for firm i to produce 1 unit of product. It is affected by firm i's productivity A_i, production technology $\alpha_{s(i)}$, the price of capital relative to labor $\frac{R}{w}$, and the input distortion which prevents firm i from reaching its profit-maximizing capital-labor ratio. Multiplying Part (a) by R gives how much firm i spends on capital for 1 unit of production. Part (b) is how much cost firm i needs to spend on capital and labor when spending 1 unit on capital. Multiplying R, Part (a), and Part (b) together gives firm i's marginal cost.

Firm i's equilibrium price P_i is its marginal cost mc_i times the markup due to its demand elasticity $\frac{\epsilon_{g(i)}}{\epsilon_{g(i)}-1}$ and the output distortion $\frac{1}{1-\tau_i^Y}$:

$$P_i = \text{mc}_i \cdot \frac{\epsilon_{g(i)}}{\epsilon_{g(i)} - 1} \cdot \frac{1}{1-\tau_i^Y} \tag{23}$$

Firm i's equilibrium market share out of its nest is:

$$\frac{P_i Y_i}{\bar{P}_{g(i)} \bar{Y}_{g(i)}} = \frac{\gamma_i^{1-\epsilon_{g(i)}}}{\sum_{i \in \mathcal{G}(g(i))} \gamma_i^{1-\epsilon_{g(i)}}} \tag{24}$$

When $\frac{1}{1-\tau_i^Y} > 1$, firm i's production is smaller than the production predicted by its marginal cost. Details on the derivations used in this section are in Appendix B.

3.3 Some special cases of the distortions

To illustrate how the input and output distortions affect the economy and the aggregate labor share, we analyze two special cases of the distortions.

Case 1: $\tau_i^I = 0$ and $\tau_i^Y = \tau \in (0, 1)$

Since input distortions are 0 and output distortions are the same for all the firms, the distortions do not affect how capital and labor are allocated across firms. The aggregate labor share after removing the distortions would increase:

$$\frac{w^* \bar{L}}{\bar{P}^* \bar{Y}^*} - \frac{w \bar{L}}{\bar{P} \bar{Y}} = \tau \sum_g \beta_g \frac{\epsilon_g - 1}{\epsilon_g} \alpha_{s(g)} > 0$$

When normalize $\bar{P}\bar{Y}$ and $\bar{P}^*\bar{Y}^*$ to 1, the wage would be $\frac{1}{1-\tau} > 1$ times higher when the distortions are removed:

$$\frac{w^*}{w} = \frac{1}{1-\tau} > 1$$

Although the distortions cause the wage to be lower, the real output does not change. This occurs because the distortions have no impact on how capital and labor are allocated, therefore no impact on the real output.

Case 2: $\tau_i^I = \tau$ and $\tau_i^Y = 0$

In this case, there is no output distortions, but the input distortions cause labor to be more expensive relative to capital if $\tau \in (-1, 0)$ and cheaper if $\tau > 0$. Therefore firm-level labor share is lower than their distortion-free values when $\tau \in (-1, 0)$ and higher if $\tau > 0$. The input distortions are the same for all the firms, which means the marginal costs of all the firms in the same nest are raised by the same factor. As a consequence, firms' market shares within nests are the same as the distortion-free ones:

$$\frac{P_i Y_i}{\bar{P}_{g(i)} \bar{Y}_{g(i)}} = \frac{A_i^{\epsilon_{g(i)}-1}}{\sum_{i \in \mathcal{G}(g(i))} A_i^{\epsilon_{g(i)}-1}} = \frac{P_i^* Y_i^*}{\bar{P}^*_{g(i)} \bar{Y}^*_{g(i)}}$$

However, because industries have different production technology α_s, labor will be reallocated across industries, which affects allocation efficiency.

When $\tau \in (-1, 0)$, for industries with a lower α_s, labor is less important, and therefore they have a smaller desire to keep using labor compared to industry with a higher α_s. The input distortions let high α_s industry to overuse labor and low α_s industry to underuse labor. Consequently, when removing the input distortions, labor would be reallocated from nests in high α_s industries to nests in low α_s industries:

$$\bar{L}^*_g - \bar{L}_g = \bar{L} \left(\frac{\beta_g \frac{\epsilon_g - 1}{\epsilon_g} \alpha_{s(g)}}{\sum_g \beta_g \frac{\epsilon_g - 1}{\epsilon_g} \alpha_{s(g)}} - \frac{\beta_g \frac{\epsilon_g - 1}{\epsilon_g} \alpha_{s(g)} \cdot \frac{1}{1+\tau\alpha_{s(g)}}}{\sum_g \beta_g \frac{\epsilon_g - 1}{\epsilon_g} \alpha_{s(g)} \cdot \frac{1}{1+\tau\alpha_{s(g)}}} \right)$$

The $\bar{L}_g$ is the $\bar{L}^*_g$ weighted by $\frac{1}{1+\tau\alpha_{s(g)}}$. The smaller the weights, the smaller the amount of labor is used in the nest under the distortions. Because the aggregate labor is fixed, the nest with the smallest weight will use more labor after the distortions are removed. Because the weights are positively correlated with α_s, nests with the smallest α_s will use more labor after the distortions are removed. Since labor allocation across nests is different from the profit-maximizing allocation, the real output is lower due to the distortions.

When $\tau > 0$, the reallocation of labor is in the opposite direction, but removing the distortions still provides efficiency gains.

Due to the reallocation of labor across industries after the distortions are removed, the aggregate labor share would increase if $\tau \in (-1, 0)$ and decrease if $\tau > 0$:

$$\frac{w^* \bar{L}}{\bar{P}^* \bar{Y}^*} - \frac{w \bar{L}}{\bar{P} \bar{Y}} = \sum_g \beta_g \frac{\epsilon_g - 1}{\epsilon_g} \alpha_{s(g)} \cdot \frac{\tau(\alpha_{s(g)} - 1)}{1 + \tau\alpha_{s(g)}} \Rightarrow \begin{cases} > 0 \text{ if } \tau \in (-1, 0) \\ < 0 \text{ if } \tau > 0 \end{cases}$$

In terms of the impact on the wage, the wage would be higher if $\tau \in (-1, 0)$ and lower if $\tau > 0$:

$$\frac{w^*}{w} = \frac{\sum_g \beta_g \frac{\epsilon_g - 1}{\epsilon_g} \alpha_{s(g)}}{\sum_g \beta_g \frac{\epsilon_g - 1}{\epsilon_g} \alpha_{s(g)} \frac{1+\tau}{1+\tau\alpha_{s(g)}}} \Rightarrow \begin{cases} > 1 \text{ if } \tau \in (-1, 0) \\ < 1 \text{ if } \tau > 0 \end{cases}$$

Summary of the two special cases

The first special case shows that the homogeneous output distortions lower the aggregate labor share because they reduce each firm's labor share but it does not change the real output since the allocation of capital and labor is unaffected. The second special case shows that different from homogeneous output distortions, homogeneous input distortions affect both the aggregate labor share and the total production. While removing the homogeneous input distortions always raises the aggregate output, the impact on the aggregate labor share depends on the sign of the homogeneous input distortions.

4 Identification and estimation

We observe firms' labor expenditure, capital, value added, total revenues, and total costs. To estimate the model, we need to make three additional assumptions:

A.1 The logarithm of firm i's total revenue-and-cost ratio equals the logarithm of its markup predicted by its demand elasticity plus a firm idiosyncratic error term η_i:

$$\ln\left(\frac{\text{revenues}_i}{\text{costs}_i}\right) = \ln\left(\frac{\epsilon_{g(i)}}{\epsilon_{g(i)}-1}\right) + \eta_i \tag{25}$$

The idiosyncratic term η_i within a nest follows a normal distribution: $\eta_i \sim \mathcal{N}(0, \sigma_{g(i)})$ for $\forall i \in \mathcal{G}(g)$

A.2 Capital rental rate is 0.2

A.3 The mode of input distortions $\ln(1+\tau_i^I)$ within an industry is 0, and the distribution is:

$$\ln(1+\tau_i^I) \sim 2\kappa_s \mathbb{1}[\tau_i^I < 0]\mathcal{N}(0, \delta_{s(i),-}) + 2(1-\kappa_s)\mathbb{1}[\tau_i^I > 0]\mathcal{N}(0, \delta_{s(i),+}) \text{ for } i \in \mathcal{S}(s) \tag{26}$$

where κ_s, $\delta_{s(i),-}$, and $\delta_{s(i),+}$ are parameters

Assumption A.1 implies that the part of revenues collected in output distortions is divided between costs and profits at the $\epsilon_{g(i)} - 1$ ratio, i.e. $\frac{\epsilon_{g(i)}-1}{\epsilon_{g(i)}}$ of the output-distortion part of the revenue is cost and $\frac{1}{\epsilon_{g(i)}}$ is profit. The decomposition of firm i's revenue becomes:

$$\underbrace{P_i Y_i}_{\text{revenue}} = \underbrace{wL_i + RK_i + \frac{\epsilon_{g(i)}-1}{\epsilon_{g(i)}}\tau_i^Y P_i Y_i}_{\text{total cost}} + \underbrace{\frac{1}{\epsilon_{g(i)}}(1-\tau_i^Y)P_i Y_i + \frac{1}{\epsilon_{g(i)}}\tau_i^Y P_i Y_i}_{\text{total profit}}$$

In Section 6.2, we discuss how our results would change if the cost-profit ratio within the output-distortion revenue is not $\epsilon_{g(i)} - 1$. In Appendix C, we explain why we choose revenue-cost ratios over other methods to quantify markups.

The parametric distributions of the idiosyncratic error term η_i imply that the revenue-cost ratios of firms inside industry s follow a mixture normal distribution:

$$\ln\left(\frac{\text{revenues}_i}{\text{costs}_i}\right) \sim \sum_{g\in\{g|\mathcal{G}(g)\subset\mathcal{S}(s)\}} \omega_g \mathcal{N}\left(\ln\left(\frac{\epsilon_g}{\epsilon_g - 1}\right), \sigma_g\right) \text{ for } i \in \mathcal{S}(s) \tag{27}$$

where $\{g|\mathcal{G}(g) \subset \mathcal{S}(s)\}$ is the set of nests inside industry s, and ω_g is the weight of nest g. The weight ω_g is the ex-ante probability that a firm in industry s belongs to nest g. The sum of all the weights in industry s equals 1, i.e. $\sum_{g\in\{g|\mathcal{G}(g)\subset\mathcal{S}(s)\}} \omega_g = 1$.

Assumption A.2 sets capital rental rate to 0.2 because Bai Chong-En et al. (2006) and Sun et al. (2011) quantify the Chinese capital return rate in 2005 as 20%.

We measure P_iY_i using firm i's value added because our production functions model labor and capital as the only production inputs. Firm i's labor and capital share, $\frac{wL_i}{P_iY_i}$ and $\frac{RK_i}{P_iY_i}$, are the observed labor and capital expenditure, wL_i and RK_i, divided by firm i's value added P_iY_i.

The parametric distribution function of the input distortions $\ln(1+\tau_i^L)$ in Assumption A.3 has mode 0 and is asymmetric. We pick this parametric distribution over normal distributions because we are concerned that the distributions of the input distortions within industries are asymmetric and that the average input distortions may not be 0. In the distribution function of Equation (26), κ_s is the share of firms in industry s whose labor-capital expenditure ratios are less than $\frac{\alpha_{s(i)}}{1-\alpha_{s(i)}}$. κ is multiplied by a factor of 2 so that the likelihood adds up to 1.

Under Assumption A.3, Equation (12) provides the distribution of $\ln\left(\frac{wL_i}{RK_i}\right) - \ln\left(\frac{\alpha_{s(i)}}{1-\alpha_{s(i)}}\right)$:

$$\begin{aligned}\ln\left(\frac{wL_i}{RK_i}\right) - \ln\left(\frac{\alpha_{s(i)}}{1-\alpha_{s(i)}}\right) \sim & 2\kappa_s \mathbb{1}\left[\frac{wL_i}{RK_i} < \frac{\alpha_{s(i)}}{1-\alpha_{s(i)}}\right] \mathcal{N}(0, \delta_{s(i),-}) \\ & + 2(1-\kappa_s)\mathbb{1}\left[\frac{wL_i}{RK_i} > \frac{\alpha_{s(i)}}{1-\alpha_{s(i)}}\right] \mathcal{N}(0, \delta_{s(i),+}) \text{ for } i \in \mathcal{S}(s)\end{aligned} \tag{28}$$

Using Equation (27), we estimate the nest structure and demand elasticity $\epsilon_{g(i)}$ by estimating a mixture normal distribution with an unknown number of components for each of the industry. Because mixture normal estimation is sensitive to outliers, we treat nests containing less than 5% of firms in their industries and nests with less than 10 firms as outliers and drop them. We use Equation (28) and the Maximum Likelihood Estimation to obtain production elasticities α_s. Equation (12) and (13) are used to calculate τ_i^L and τ_i^Y.

We calculate firm productivity following Hsieh and Klenow (2009) because, like them, we only care about firms' productivity relative to other firms in the same nest:

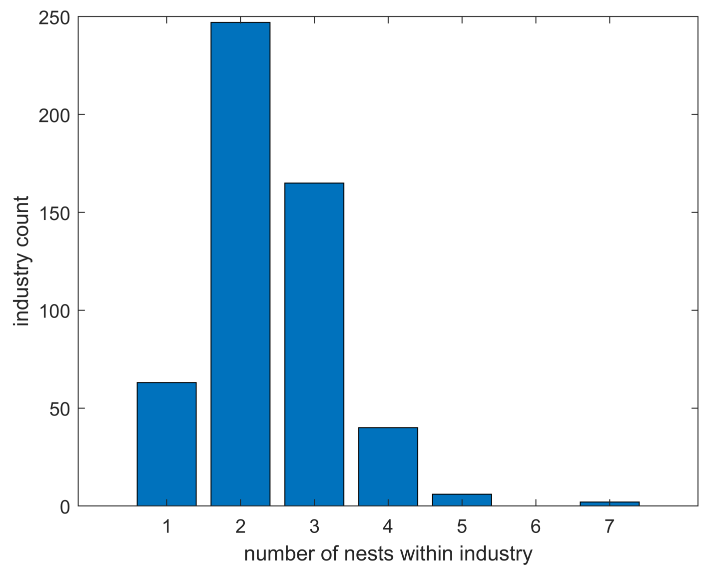
$$A_i \propto \frac{(P_i Y_i)^{\frac{\epsilon_g}{\epsilon_g - 1}}}{K_i^{1-\alpha_{s(i)}} L_i^{\alpha_{s(i)}}} \text{ for } i \in \mathcal{G}(g) \tag{29}$$

5 Results

In this section, we will first present the estimated nest structure, demand elasticities, production elasticities, and distortions. We then use these parameters to predict how and why the aggregate labor share would change when removing the distortions.

5.1 Estimated parameters and distortions

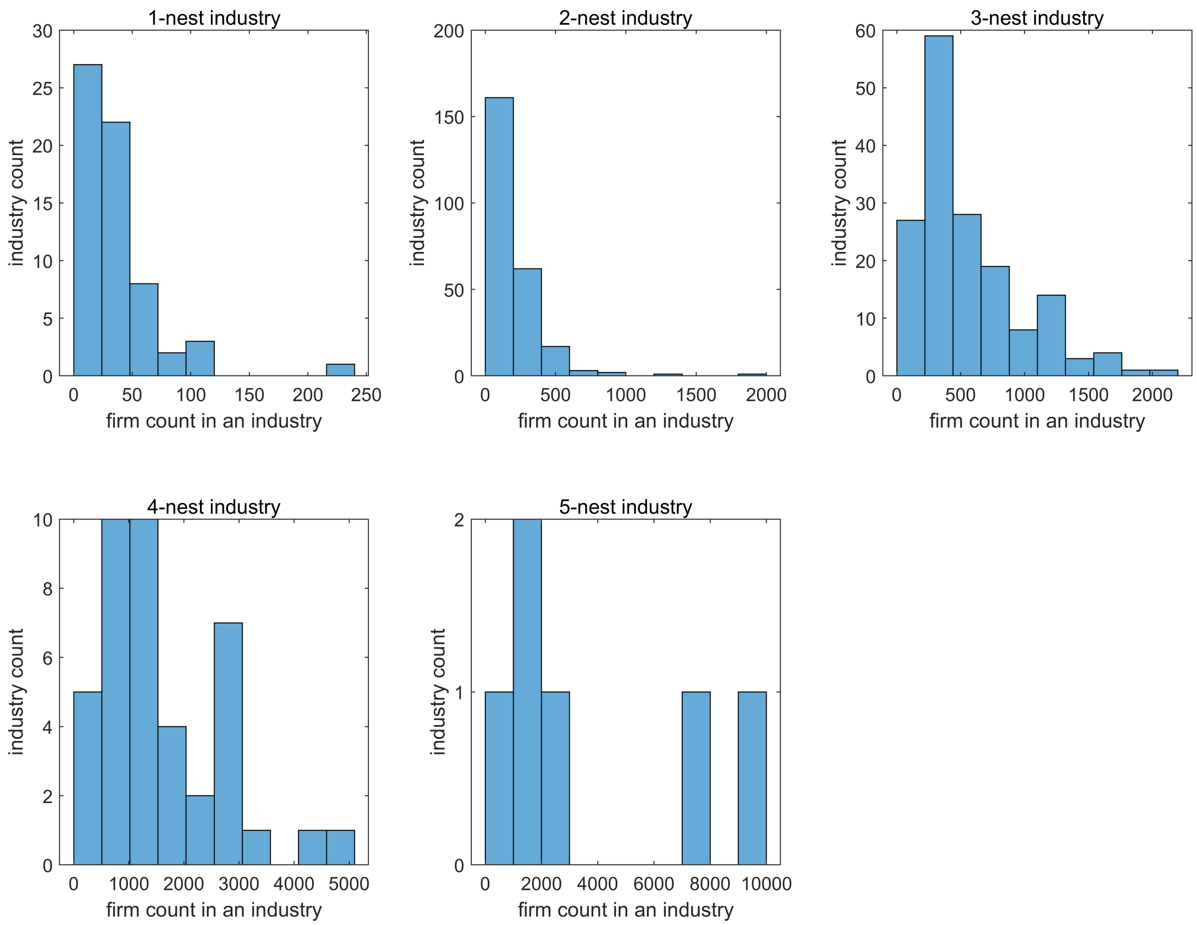
FIGURE 6: Distribution of industries by their number of nests



We plot the distribution of industries by their estimated number of nests in Figure 6 and the distribution of industry sizes for industries with the same number of nests in Figure 7. Most of the industries are estimated to have either 2 or 3 nests. The industry with the largest number of nests is estimated to have 7 nests (Figure 6). Industries with more nests generally contain more firms (Figure 7). In total, we estimate that there are 1256 nests in the economy.

The first three rows of Table 2 are the summary statistics of firms' demand elasticities. The first row is unweighted, the second row is weighted by firms' costs, and the third row is

Figure 7: Distribution of industry firm counts



weighted by firms' revenues. The average demand elasticities is 9.34 and 10.19 if weighted by revenues and 10.69 if weighted by costs. To compare our estimated demand elasticities to the estimates in existing studies, we calculate the weighted average markups predicted by the demand elasticities, i.e. $\frac{\epsilon_g}{\epsilon_g - 1}$. We use two types of weights, firm costs and revenues. Our cost-weighted average markup is 1.15, which coincides with the 1.15 cost-weighted average markups in Edmond et al. (2019). It is also consistent with Baqaee and Farhi (2020)'s estimate when using the method developed by De Loecker and Warzynski (2012). De Loecker and Warzynski (2012) themselfves estimate average markups to be between 1.10 and 1.28, a range contains our estimate. In terms of sales-weighted average markup, ours is 1.18, which is below the estimate from De Loecker et al. (2020) whose sales-weighted average markups are 1.20 in 1980 and 1.60 in 2012. Our median markup is 1.12 when weighted by costs and 1.14 when weighted by revenues, lower than the 1.30 median by Feenstra and Weinstein (2017). All these studies mentioned so far use American data. Our estimates are either the same or a bit lower than their estimates. Compared to firms from developing countries, our 1.15 average is higher than the 1.12 average markups found by Peters (2020) using Indonesian data.

Table 2: Summary statistics of selected estimated parameters

	Mean	St. Dev.	Pctl(10)	Pctl(25)	Median	Pctl(75)	Pctl(90)
ϵ_g	9.34	5.21	4.07	5.68	8.30	11.29	14.92
ϵ_g (cost)[1]	10.69	6.22	4.69	6.72	9.44	13.34	17.21
ϵ_g (revenue)[2]	10.19	6.11	4.33	6.15	9.07	12.46	16.51
α_s	0.78	0.10	0.67	0.74	0.80	0.85	0.90
α_s (cost)[1]	0.77	0.09	0.66	0.73	0.78	0.84	0.89
α_s (revenue)[2]	0.77	0.10	0.65	0.73	0.78	0.84	0.89

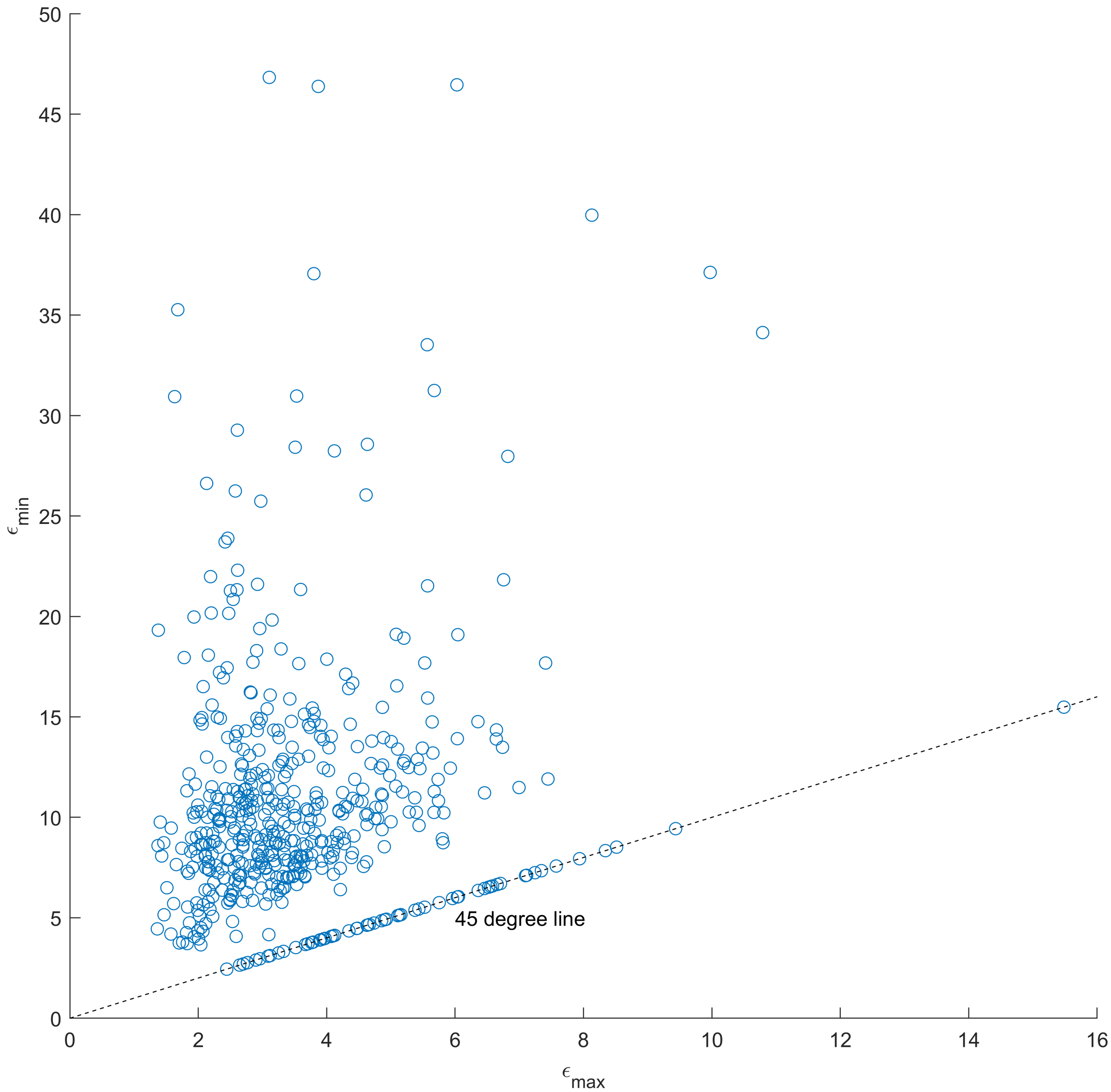
[1] The distribution is weighted by firms' costs.
[2] The distribution is weighted by firms' revenues.

The last three rows of Table 2 record the summary statistics of labor production elasticities, α_s. Similar to the first three rows, the forth row is unweighted, the fifth row is weighted by costs, and the sixth row is weighted by revenues across all the firms.

We plot an industry's largest demand elasticity against its smallest demand elasticity in Figure 8 to display the dispersion of demand elasticities within industries. Each point represents an industry. Industries with one nest lie on the 45 degree line because their maximum and minimum demand elasticities are the same.

In general, we expect competition to be stronger and hence demand elasticities to be higher in markets with more firms. Figure 9 shows that larger nests in an industry tend to

Figure 8: Maximal and minimal estimated demand elasticities within industries



have larger demand elasticities ϵ_g compared to smaller nests in the same industry. The nest size is a nest's firm count. Since different industries have different demand elasticities due to industry-specific features, we normalize the nest size by first dividing it by its industry's firm count and then subtracting the inverse of the number of nest in the industry. We normalize the demand elasticities by subtracting from them the industry average demand elasticity. Figure 9 shows that nest whose size is above the industry average tend to have above average demand elasticity and nest whose size is below the industry average tend to have below average demand elasticity, i.e. most of industries fall in the upper-right and the lower-left part of the plot.

Since there are in total 523 industries and 1256 nests, it is difficult to report all the estimated demand elasticities. So we calculate the average demand elasticities at the 2 digit industry level and report here a selection of the 2-digit industries. Oil and gas extraction, tobacco, and pharmaceutical industries have low demand elasticities. This is in line with the intuition that these industries usually have strong market power. The second observation is that although sometimes larger industries have higher demand elasticities, such as the textile industry, it is not always the case. For example, the pharmaceutical industry contains 1.5 times more firms than the rubber products industry but its demand elasticity is about 1/2 of that of the rubber products industry.

TABLE 3: Average demand elasticities at the 2-digit industry level for selected industries

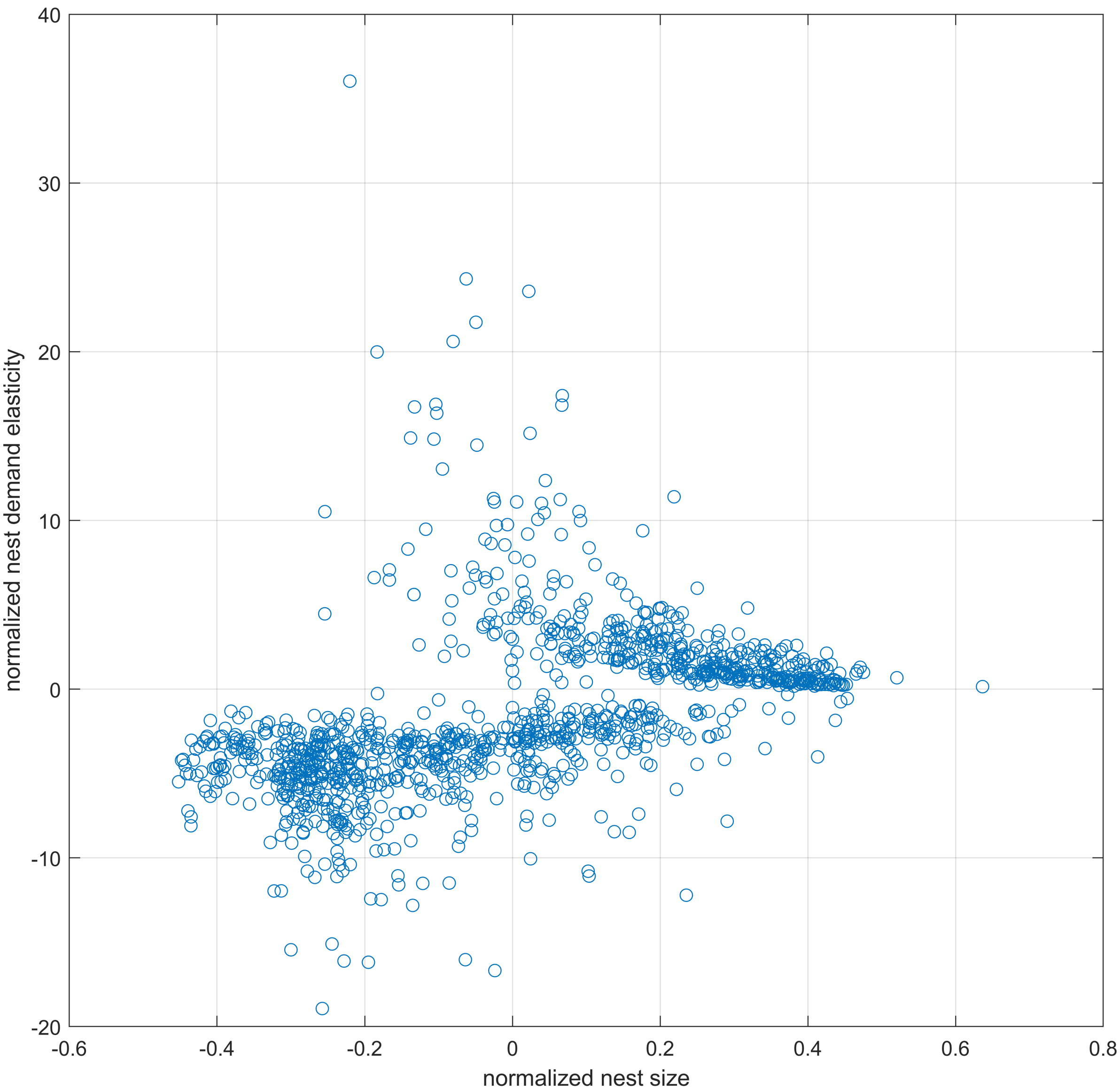
Industry	ϵ	**Firm counts**
Oil and gas extraction	2.80	33
Agricultural and Sideline Food Processing	7.90	12273
Tobacco	2.78	102
Textile	10.05	20183
Pharmaceutical	2.99	4267
Rubber products	6.67	2699

5.2 Estimated distortions

Figure 10 presents the distributions of firms' input and output distortion. The distributions of input distortions is roughly symmetric around 0 even though we do not assume 0 median or mean for input distortions. Output distortions are skewed to the left. The distribution of input and output distortions implies that about half of the firms underuse labor relative to capital but most of the firms are too small.

Figure 11 plots the distribution of input and output distortions by ownership types. State owned enterprises (SOEs) tend to be bigger than their profit-maximizing sizes (positive

Figure 9: Nest sizes and demand elasticities



Notes: Normalized nest size is $\frac{N_g}{N_{s(g)}} - \frac{1}{G_{s(g)}}$, where N_g and $N_{s(g)}$ are the firm counts of nest g and industry $s(g)$, and $G_{s(g)}$ is industry $s(g)$'s nest count. Normalized nest demand elasticity is $\epsilon_g - \bar{\epsilon}_{s(g)}$, where $\bar{\epsilon}_{s(g)}$ is the average demand elasticity of firms in industry $s(g)$, $\bar{\epsilon}_{s(g)} = \frac{1}{N_s} \sum_{i \in \mathcal{S}(s)} \epsilon_i$. We remove industries with only 1 nest.

Figure 10: Estimated distortions $\log(1-\tau_i^Y)$ and $\log(1+\tau_i^I)$

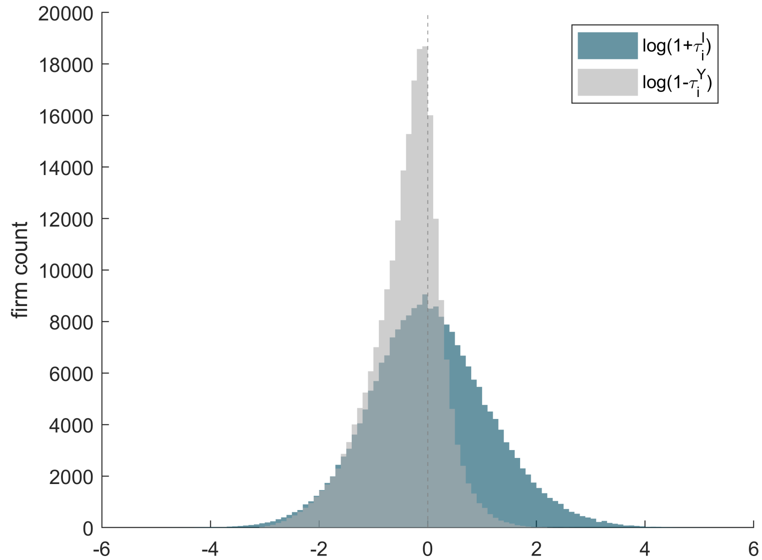


Figure 11: Estimated distortions $\log(1-\tau_i^Y)$ and $\log(1+\tau_i^I)$ by ownership

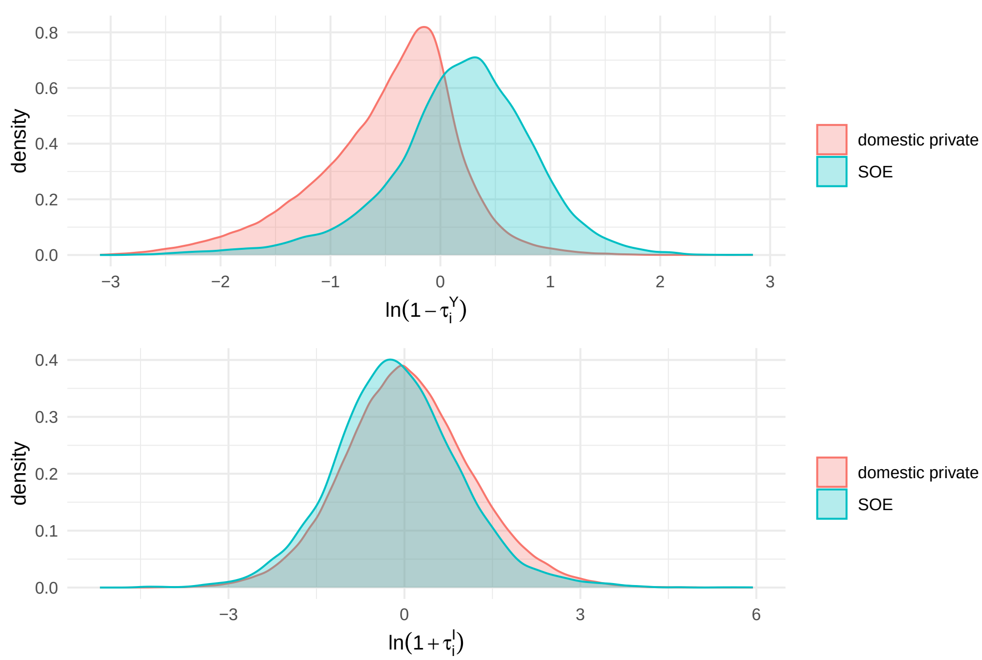


FIGURE 12: Estimated distortions $\log(1-\tau_i^Y)$ and $\log(1+\tau_i^I)$ by export status

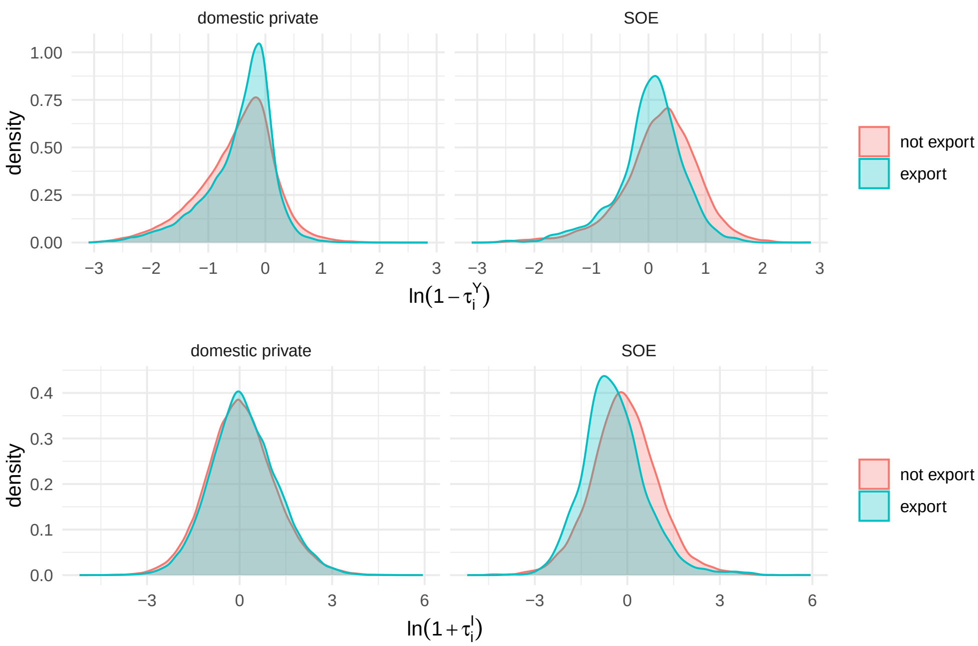


FIGURE 13: Estimated distortions $\log(1-\tau_i^Y)$ and $\log(1+\tau_i^I)$ by patent status

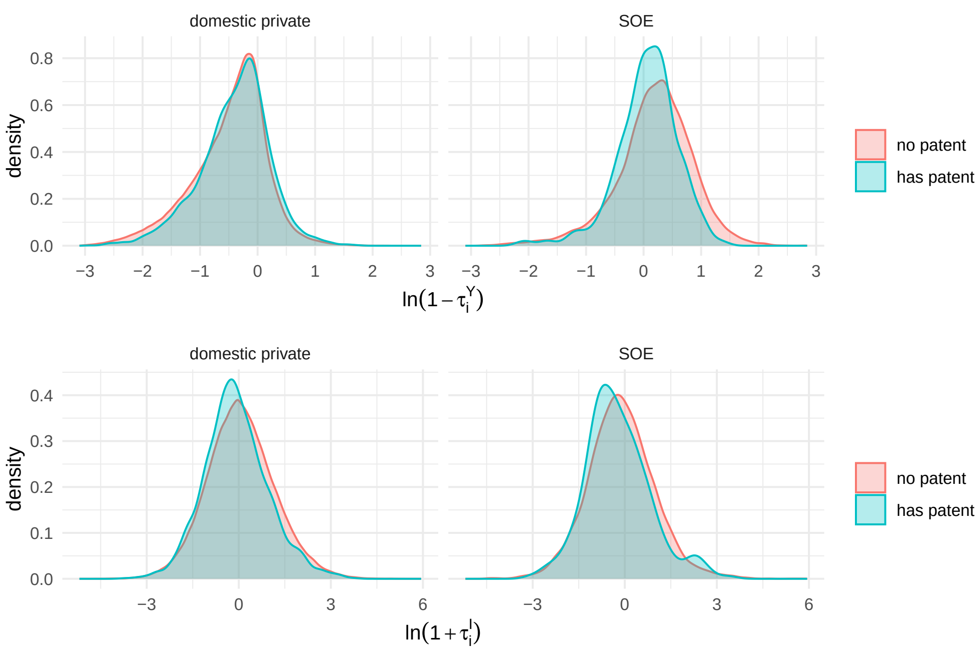


FIGURE 14: Estimated distortions $\log(1-\tau_i^Y)$ and $\log(1+\tau_i^I)$ by loan status

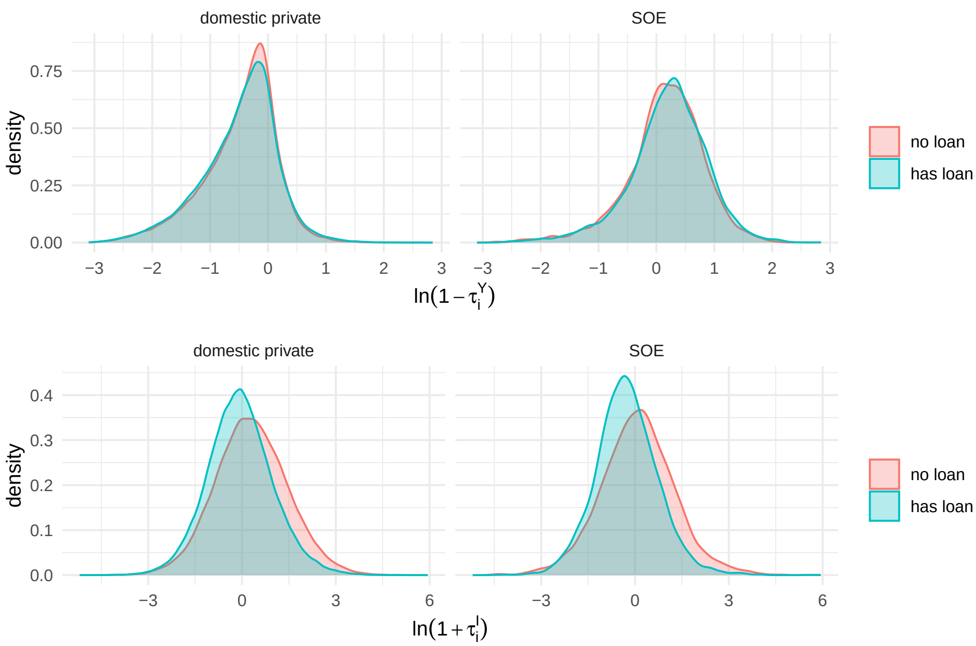
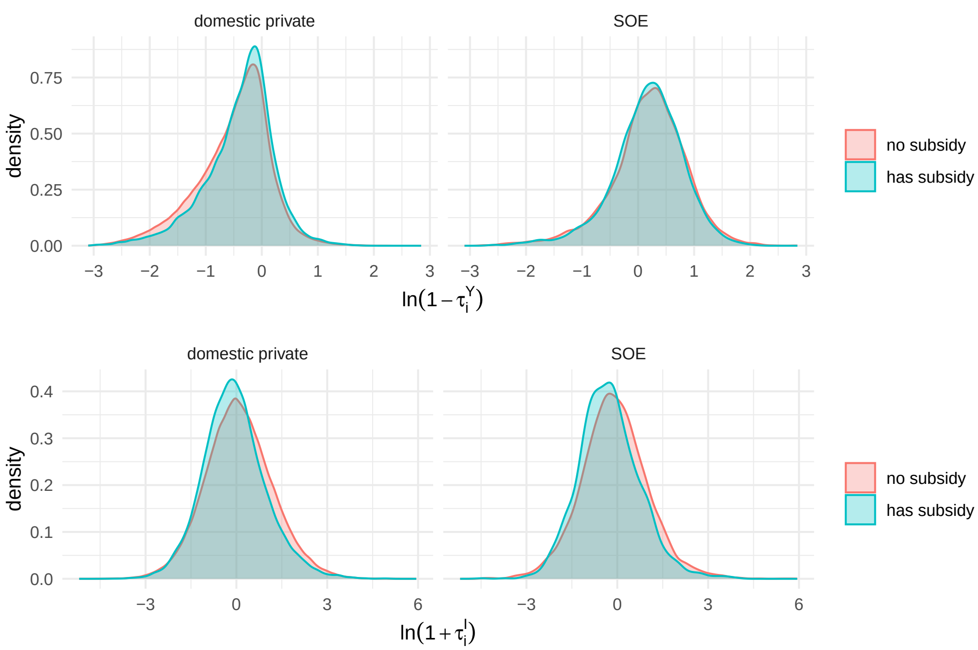


FIGURE 15: Estimated distortions $\log(1-\tau_i^Y)$ and $\log(1+\tau_i^I)$ by subsidy status



$ln(1-\tau_i^Y)$), whereas domestic private firms tend to be smaller than their profit-maximizing sizes (negative $ln(1-\tau_i^Y)$). SOEs are slightly more likely to overuse capital relative to labor (negative $ln(1+\tau_i^I)$) compared to domestic private firms. This is consistent with the idea that SOEs tend to grow excessively large due to political interventions and cheap capital credits and that domestic private firms are generally restrained from reaching their full potentials due to the lack of access to the financial market and other frictions.

Within ownership, SOEs that export or have patents are less likely to be excessively large than non-export SOEs, but they tend to have lower labor-capital ratios (Figure 12 and 13). This suggests that SOEs more exposed to market competition, i.e. exporting SOEs, and SOEs that has more competency, i.e. SOEs with patents, face smaller size distortions. The impact of loans and subsidies are overall small. They do not affect the distribution of output distortions, but firms with loans or subsidies tend to have slightly negative input distortions, i.e. slightly lower labor-capital ratios (Figure 14 and Figure 15).

5.3 Predicted impact of removing the distortions

TABLE 4: Observed and predicted aggregate labor shares (%)

	with input and output distortions	with only input distortions	without distortions
wL/PY	42	65.03	65.88

The aggregate labor share with input and output distortions, or in other words the one observed in the data, is 42%. Removing all the distortions raises the aggregate labor share to 66%, i.e. an increase of 24 percentage points. This change is primarily driven by the output distortions because removing the output distortions alone raise the aggregate labor share to 65%. Table 4 summarizes these predicted aggregate labor shares.

The distributions of distortions in Figure 10 imply that half of the firms would reduce their labor demand after removing the input distortion but the other half would raise their labor demand. Removing the output distortions would raises 70% of the firms' labor demand.

Figure 16 presents the distribution of changes in firm-level labor shares. 76% the firms' labor shares increase after removing the distortions. Comparing the distribution of firms' input and output distortions in Figure 10 to firms' labor share changes in Figure 16, we interpret that the predicted firm-level labor share changes are also mainly driven by the output distortions.

Figure 17 is the distribution of changes in firm-level labor-capital expenditure ratios. It

Figure 16: Predicted changes in firm-level labor shares

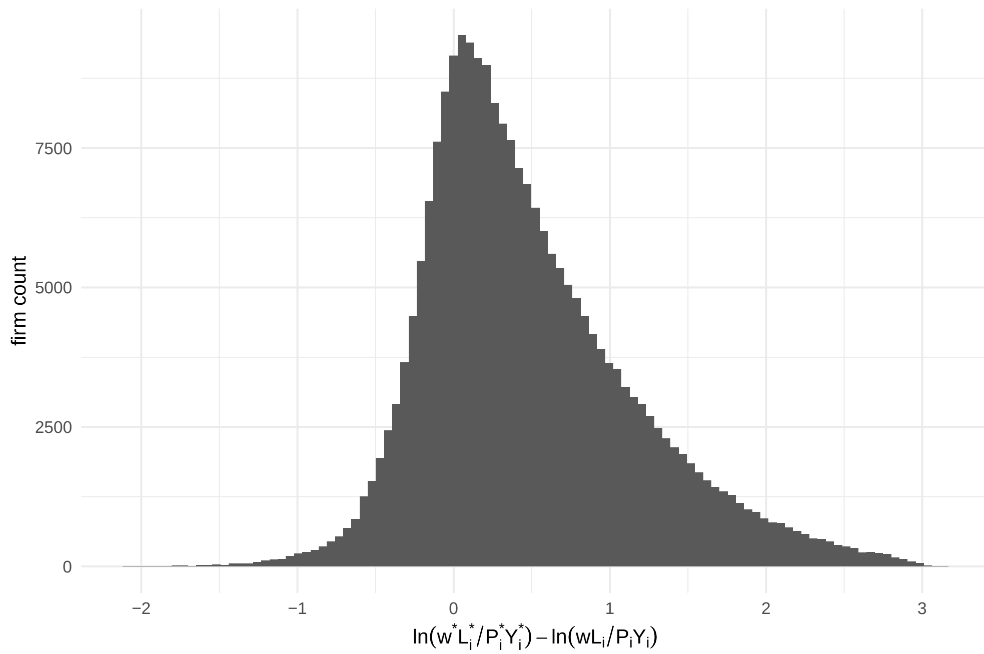
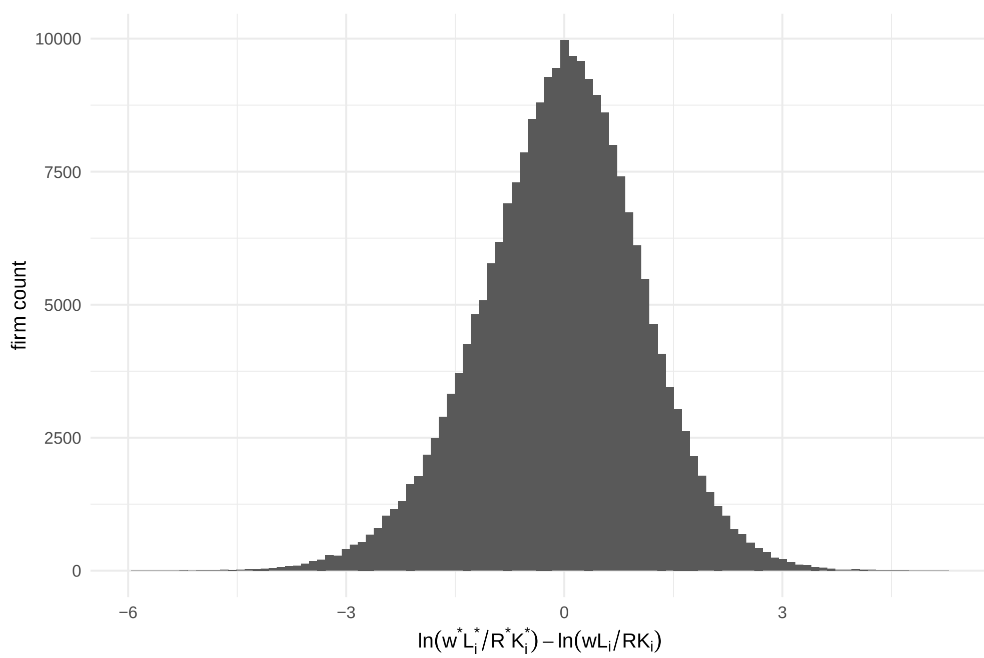


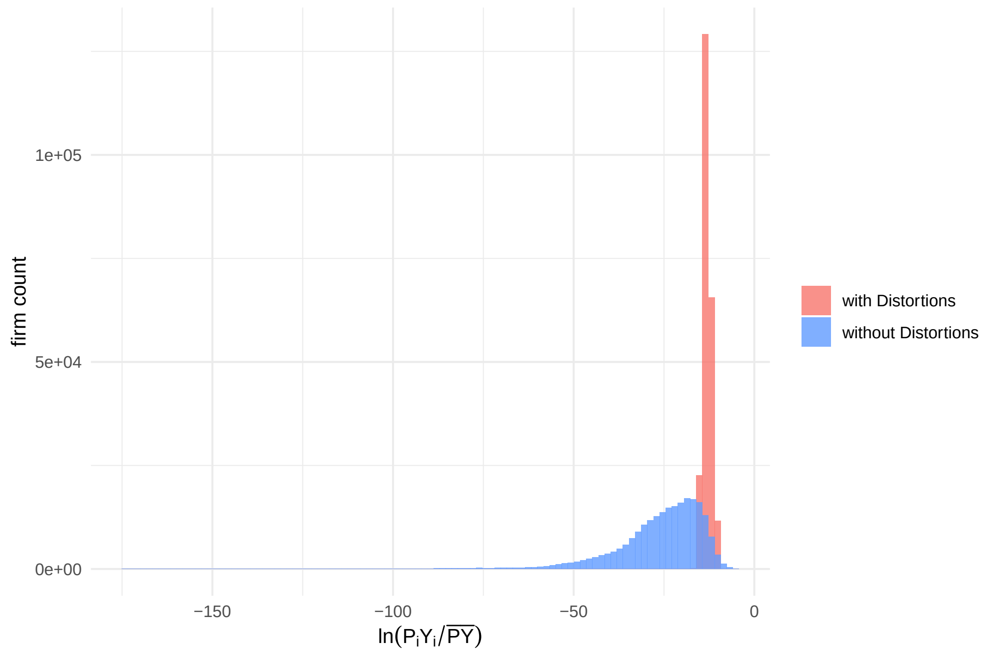
Figure 17: Predicted changes in firm-level expenditure ratios of labor and capital



is the same as the distribution of the input distortions in Figure 10 because the changes in firm-level expenditure ratios equal input distortions.

The reallocation across firms increases the dispersion of firm sizes. More specifically, it causes some firms to grower bigger and others to shrink as shown in the distribution of firm sizes with and without the distortions in Figure 18. As a consequence, market concentration increases. Figure 19 plots the distribution of industry Herfindahl–Hirschman index. Removing the distortions shifts the distribution of industry Herfindahl–Hirschman index to the right. Some industry's Herfindahl–Hirschman index becomes almost 10000, which implies that almost the entire industry is served by one firm.

FIGURE 18: Distribution of firm shares with and without distortions



Due to the unitary elasticity of substitution across industries, removing the distortions does not alter industry shares $\frac{\bar{P}_s \bar{Y}_s}{\bar{P}\bar{Y}}$. However, removing the distortions alters industry-level labor demand and consequently changes the industry-level labor use. We plot the distribution of the changes in industry-level labor use in Figure 20. The distribution has a positive skewness, indicating that more than half of the industries increase their labor use. However, among industries with large labor use change, i.e. $|\ln(\bar{L}_s^*) - \ln(\bar{L}_s)| > 0.72$, more industries reduce their labor use.

Comparing the changes in industry-level labor use in Figure 20 to the changes in industry-level labor share in Figure 21, we can see that although only half of the industries' labor uses increase, the majority of the industries increase their labor shares. Industry-level labor shares increase after removing the distortions when the distortions force these industries to be too small and hinder the optimal allocation of capital and labor within the industries. Most of the industries' labor demand in our data are suppressed by the distortions. However, because the aggregate labor is fixed, whether an industry's labor use increases depends whether its demand for labor increase disproportionally more compared to other industries. Therefore,

FIGURE 19: Industry Herfindahl–Hirschman index with and without distortions

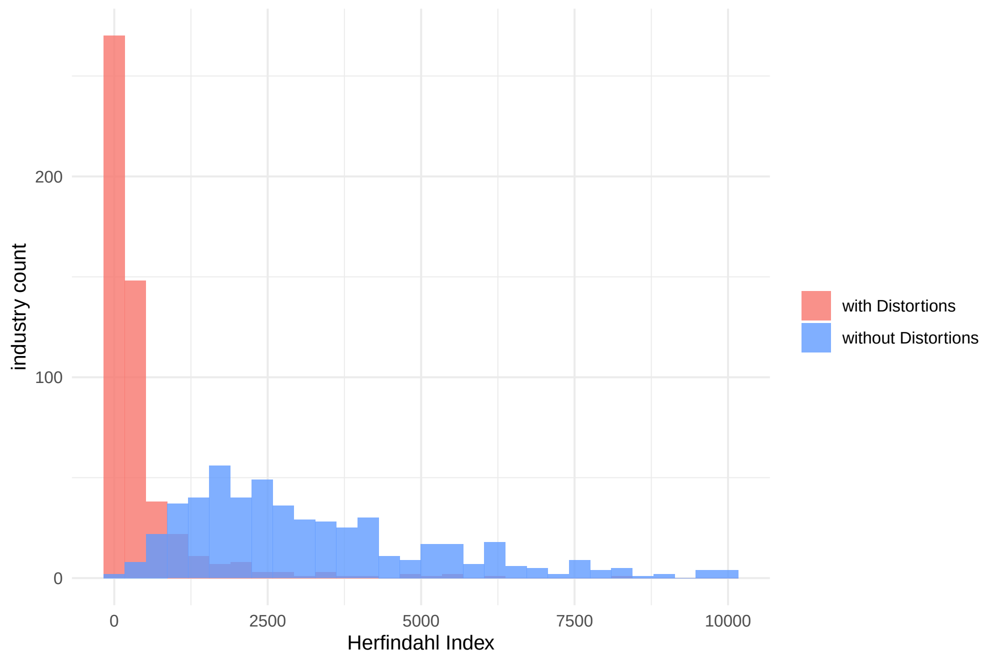
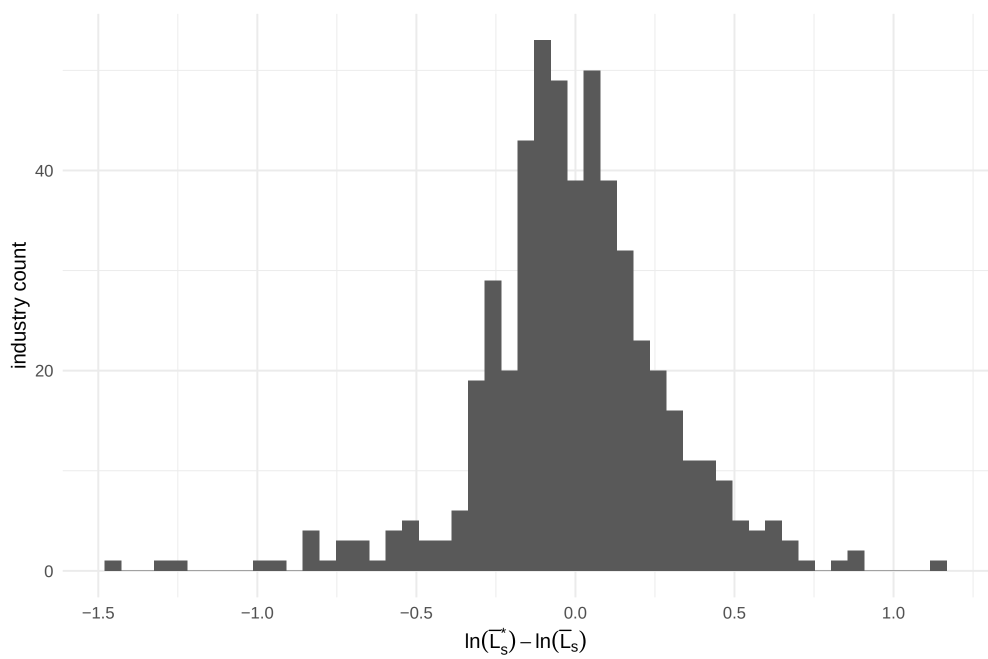


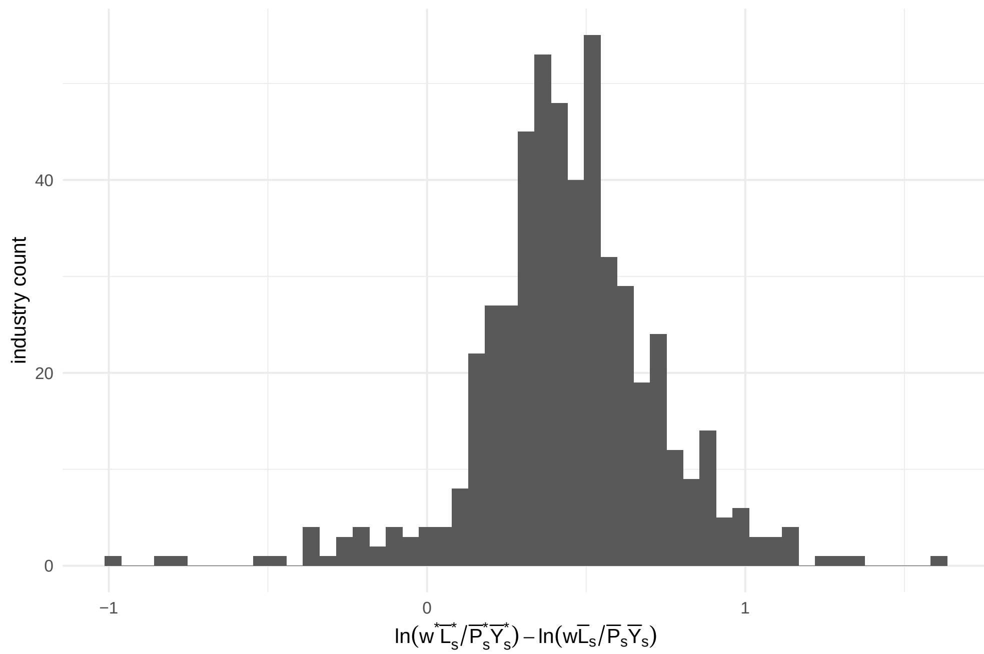
FIGURE 20: Changes in industry-level labor use



only half of the industries manage to increase their labor use.

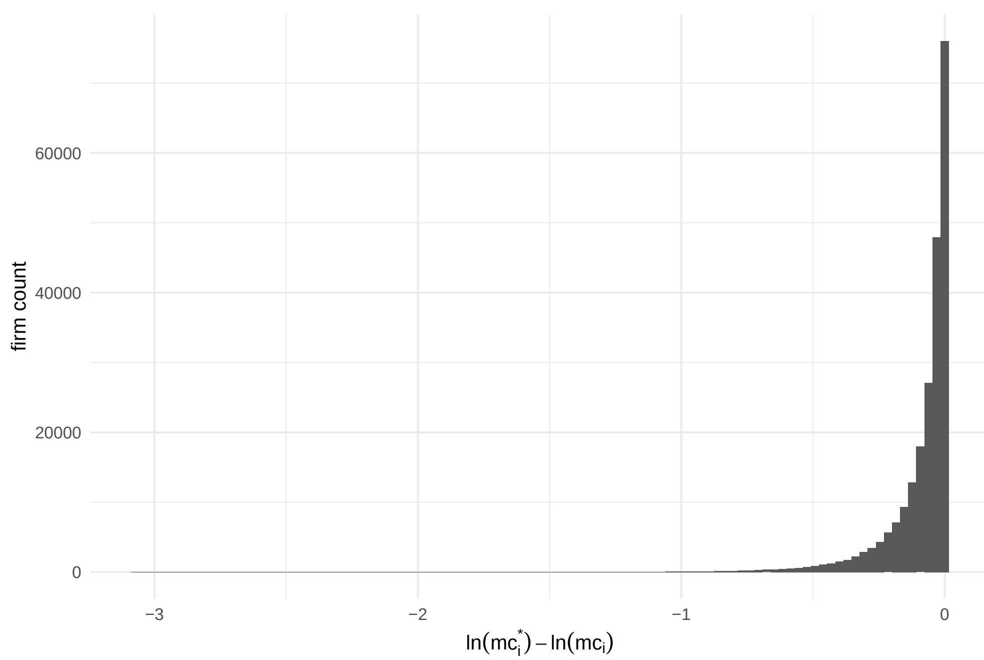
Removing input distortions allows firms to reach their profit-maximizing labor-capital ratios and consequently reduces firms' marginal costs. Figure 22 plots the predicted changes in firms' marginal costs from removing the input distortions holding wage and capital rental

FIGURE 21: Changes in industry-level labor shares



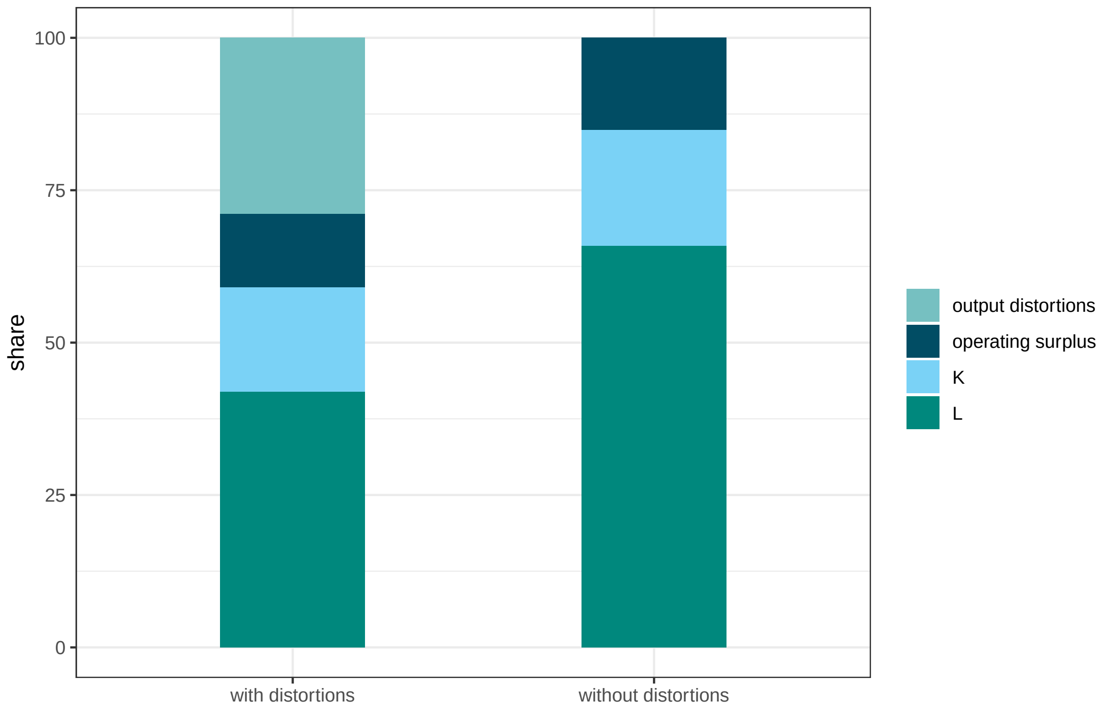
rate fixed. The majority of firms' marginal costs reduce by less than 50% but the distribution has a long left tail, indicating a small share of firms experience 95% ($1 - e^{-3} = 95\%$) reduction in their marginal costs.

FIGURE 22: Changes in firm marginal costs



We conclude the result section by the decomposition of the aggregate revenue (Figure 23). With distortions, labor collects 42% of the aggregate income, capital 17%, output distortions 29%, and firms' operating surplus 12%. Firms' operating surplus is the part of the profits due to markups predicted by demand elasticities. After removing the distortions, labor receives 66%, capital 19%, and operating surplus 15%, which is also the profit share since there is no distortion any more.

FIGURE 23: Decomposition of the aggregate revenue



6 Robustness

In this section, we discuss how to interpret our results if some of our assumptions are violated.

6.1 Heterogeneous ratios between observed and unobserved labor compensation

If the ratios between observed and unobserved labor compensation vary across firms, correcting labor shares using a common factor would treat the variation in unobserved labor shares as the input distortions. If the variation in unobserved labor shares move in the same direction as the input distortions, we overestimate the input distortions. Because our results show that the main impact on the aggregate labor share is from the output distortions, the heterogeneous ratios between observed and unobserved labor compensation should have a marginal impact on the aggregate labor share change. However, if the variation in unobserved labor shares move in the same direction as the input distortions, we would underestimate the input distortions and possibly the impact of the input distortions on the aggregate labor share.

6.2 Profits from the output-distortion part of firm revenues are not $\frac{1}{\epsilon_g}\tau_i^Y P_i Y_i$

In the main results, we assume that $\frac{1}{\epsilon_g}$ of firms' output-distortion revenue is profits. This means that firms' revenue-cost ratios are not affected by the output distortions:

$$\ln\left(\frac{\text{cost}_i}{\text{revenue}_i}\right) = \ln\left(\frac{\epsilon_{g(i)}-1}{\epsilon_{g(i)}}\right) + \eta_i$$

When this is not the case, our estimated demand elasticities can be contaminated by output distortions. In the extreme case where the output-distortion part of revenues are all profits, the logarithm of firms' revenue-cost ratios are:

$$\ln\left(\frac{\text{cost}_i}{\text{revenue}_i}\right) = \ln\left(\frac{\epsilon_{g(i)}-1}{\epsilon_{g(i)}}\right) + \ln(1-\tau_i^Y) + \eta_i$$

When the output-distortion part of revenues are all costs, the logarithm of firms' revenue-cost ratios are:

$$\ln\left(\frac{\text{cost}_i}{\text{revenue}_i}\right) = \ln\left(1-(1-\tau_i^Y)\frac{1}{\epsilon_{g(i)}}\right) + \eta_i$$

In both scenarios, we expect to see the variation in the revenue-cost ratios to be the same as or very close to the variation in the cost shares of labor and capital because both contains the variation from demand elasticities and from output distortions. However, our data shows that the standard deviation of firms' cost shares of labor and capital is 0.42 and the standard deviation of firms cost-revenue ratios is 0.12, i.e. the cost shares are more volatile than the cost-revenue ratios. We interpret this as a suggestion that the data is better described by assuming the cost-revenue ratios do not contain the output distortions $1-\tau_i^Y$.

6.3 Heterogeneity in firms' capital-labor expenditure shares reflects heterogeneous technology

Our paper argues for the importance of taking into account firm heterogeneity when evaluating the impact of removing the distortions, but we assume away production technology difference within industry apart from firm-specific productivity A_i. If the production technology differs across firms within industries, i.e. α_i is firm-specific, both the input distortions and this technology heterogeneity will contribute to the dispersion of firm-level expenditure ratios between labor and capital:

$$\ln\left(\frac{wL_i}{RK_i}\right) = \ln(1+\tau_i^I) + \ln\left(\frac{\alpha_i}{1-\alpha_i}\right)$$

If the input distortions τ_i^I are uncorrelated with technology α_i, the variation of firms' labor-capital expenditure shares is the sum of the variation of input distortions and technology:

$$\text{Var}\left(\frac{wL_i}{RK_i}\right) = \text{Var}(1+\tau_i^I) + \text{Var}\left(\frac{\alpha_i}{1-\alpha_i}\right)$$

In this case, we overestimate the input distortions because we attribute some of the technology heterogeneity to input distortions. However, we argue that it is unlikely to change the estimated impact on the aggregate labor share because technology heterogeneity does not affect the estimated output distortions, and our predicted labor share change is primarily driven by the output distortions.

7 Conclusion

Reallocating capital and labor to increase the total production and to ensure that labor receives a fair share of the aggregate income is important for both economists and policymakers. Grossman and Oberfield (2022) observe that there is a radical but currently popular concern that the aggregate labor share in the national income might fall to zero due to market concentration and technological progress, such as further automation and the adoption of artificial intelligence. The problem of how to pursue higher productivity and to avoid a falling labor share is difficult when increasing productivity lowers labor share. Our estimates show that Chinese firms in 2005 face distortions that limit their growth and removing these distortions leads to the win-win outcome where both the total production and the aggregate labor share increase. This implies that policymakers of countries restrained by similar allocation distortions should focus on improving allocation efficiency and support their firms' growth.

Our quantified labor-share increase is under the assumption that the aggregate labor and capital is fixed. Removing the distortions unleashes firms' demand for labor and pushes up wage by 57% since firms have to fight for labor. This should be interpreted as a short-term impact because, in the long run, the 57% wage increase will attract a huge inflow of labor, especially in the Chinese context where the agriculture sector employs 45% of national aggregate workers but its wage is only 50% of the one in the manufacturing sector and even lower than the mining sector and the public utilities sector (Chinese Annual Yearbook). The socioeconomic impact of a worker migration of this magnitude would not be trivial as suggested by Pellegrina and Sotelo (2021). How to prepare society for the migration should be considered by economists and policymakers.

References

Acemoglu, D. (2003): "Labor- and Capital-Augmenting Technical Change," *Journal of the European Economic Association*, 1, 1–37, publisher: Oxford University Press.

Acemoglu, D. and P. Restrepo (2018): "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, 108, 1488–1542.

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.

Bai, C.-E. and Z. Qian (2010): "The factor income distribution in China: 1978–2007," *China Economic Review*, 21, 650–670.

Bai Chong-En, Chang-Tai Hsieh, and Yingyi Qian (2006): "The Return to Capital in China," *Becker Friedman Institute.*

Baqaee, D. R. and E. Farhi (2020): "Productivity and Misallocation in General Equilibrium," *The Quarterly Journal of Economics*, 135, 105–163.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841.

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.

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

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The Rise of Market Power and the Macroeconomic Implications," *The Quarterly Journal of Economics*, 135, 561–644.

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

Edmond, C., V. Midrigan, and D. Y. Xu (2019): "How Costly Are Markups," *working paper.*

Feenstra, R. C. and D. E. Weinstein (2017): "Globalization, Markups, and US Welfare," *Journal of Political Economy*, 125, 1040–1074, publisher: The University of Chicago

Press.

Grossman, G. M. and E. Oberfield (2022): "The Elusive Explanation for the Declining Labor Share," *Annual Review of Economics*, 14, 93–124, publisher: Annual Reviews.

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.

Jones, B. F. and X. Liu (2024): "A Framework for Economic Growth with Capital-Embodied Technical Change," *American Economic Review*, 114, 1448–87.

Klette, T. J. and Z. Griliches (1996): "The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous," *Journal of Applied Econometrics*, 11, 343–361.

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

Liang, Y. (2023): "Misallocation and markups: Evidence from Indian manufacturing," *Review of Economic Dynamics*, 51, 161–176.

Liu, E. (2019): "Industrial Policies in Production Networks," *The Quarterly Journal of Economics*, 134, 1883–1948.

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

Pellegrina, H. and S. Sotelo (2021): "Migration, Specialization, and Trade: Evidence from Brazil's March to the West," Tech. Rep. w28421, National Bureau of Economic Research, Cambridge, MA.

Peters, M. (2020): "Heterogeneous Markups, Growth, and Endogenous Misallocation," *Econometrica*, 88, 2037–2073.

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

Ridder, M. D., B. Grassi, and G. Morzenti (2021): "The Hitchhiker's Guide to Markup Estimation," Working Papers 677, IGIER , Bocconi University.

Sun, W., X. Yang, and G. Xiao (2011): "Understanding China's High Investment Rate and FDI Levels: A Comparative Analysis of the Return to Capital in China, the United States, and Japan," *United States International Trade Commission, Journal of International Commerce and Economics*.

Zhang, X. and J. Xia (2022): "Misallocation under Heterogeneous Markups," *working paper*.

Appendix

A Data

We drop observations with negative value added, negative wage expenditures, negative capital, negative total assets, negative account receivable, negative total debts, negative long-term debts, negative account payable, negative exports, negative sales, and negative costs. We also drop observations whose account receivable is larger than total assets, total debts larger than total assets, account payable larger than liquid debts, and profits larger than sales. If a firm's costs are missing but its sales and profits are observed, then its costs are sales minus profits. The survey reports firms' net value of capital and investment. To calculate depreciated net value of capital, we use perpetual annuity method following Brandt et al. (2012).

TABLE 5: Summary statistics unweighted across firms

	Mean	Standard Deviation	10th Percentile	1st Quartile	Median	3rd Quartile	90th Percentile
labor share	0.32	0.32	0.06	0.12	0.23	0.42	0.65
adj. labor share	0.49	0.26	0.16	0.28	0.47	0.68	0.85
capital share	0.19	0.30	0.02	0.05	0.10	0.22	0.43
cost share	0.68	0.42	0.24	0.40	0.64	0.87	1.13
cost/revenue	0.85	0.12	0.70	0.80	0.88	0.93	0.96

Notes: Total number of firms is 229,282. Capital rental rate is assumed 0.2. Cost share is adjusted labor share plus capital share.

B Derivation

Solving the firms' problem in Equation (8) gives firm i's marginal cost mc_i and price P_i:

$$\text{mc}_i = R \cdot A_i^{-1} \left[\frac{R}{w} \cdot \frac{\alpha_{s(i)}}{1-\alpha_{s(i)}} \cdot (1+\tau_i^I) \right]^{-\alpha_{s(i)}} \cdot \left[1 + \frac{\alpha_{s(i)}}{1-\alpha_{s(i)}} (1+\tau_i^I) \right] \tag{30}$$

$$P_i = \text{mc}_i \cdot \frac{\epsilon_{g(i)}}{\epsilon_{g(i)}-1} \cdot \frac{1}{1-\tau_i^Y} \tag{31}$$

Combining Equation (17), (30), and (31) gives firm i's market share out of nest $g(i)$:

$$\frac{P_i Y_i}{\bar{P}_{g(i)} \bar{Y}_{g(i)}} = \frac{\gamma_i^{1-\epsilon_{g(i)}}}{\sum_{i \in \mathcal{G}(g(i))} \gamma_i^{1-\epsilon_{g(i)}}} \tag{32}$$

where

$$\gamma_i \equiv \frac{1}{1-\tau_i^Y} \cdot \frac{1}{A_i} \left[(1+\tau_i^I)^{-\alpha_{s(i)}} \cdot (1-\alpha_{s(i)}) + (1+\tau_i^I)^{1-\alpha_{s(i)}} \cdot \alpha_{s(i)} \right] \tag{33}$$

Substitute Equation (16) and (32) for $\frac{\bar{P}_g \bar{Y}_g}{\bar{P}\bar{Y}}$ and $\frac{P_i Y_i}{\bar{P}_{g(i)} \bar{Y}_{g(i)}}$ gives Equation (18):

$$\frac{w\bar{L}}{\bar{P}\bar{Y}} = \sum_g \sum_{i \in \mathcal{G}(g)} \beta_g \cdot \underbrace{(1-\tau_i^Y) \frac{\epsilon_{g(i)} - 1}{\epsilon_{g(i)}} \cdot \frac{\alpha_{s(i)}(1+\tau_i^I)}{1+\alpha_{s(i)}\tau_i^I}}_{\text{firm-level labor shares}} \cdot \underbrace{\frac{\gamma_i^{1-\epsilon_{g(i)}}}{\sum_{i \in \mathcal{G}(g(i))} \gamma_i^{1-\epsilon_{g(i)}}}}_{\text{firm-level market shares}}$$

C Inferring markups using revenue-cost ratio versus other methods

Inferring markups without observed prices, physical production, and physical inputs is difficult. Generally, there are three methods for estimating markups: the demand approach, the production approach, and the accounting approach. Developed by Berry et al. (1995), the demand approach models consumers' choices among products and infers markups from parameters in consumers' utility functions. This method requires product prices, sales in units of products, and some observed characteristics of the products. The production approach measures markups as the ratio of production elasticities to cost share of a variable input (De Loecker and Warzynski (2012)). Although it does not require prices, applying it to markets with heterogeneous markups and heterogeneous production functions creates various problems when physical production and physical inputs are replaced by revenue production and input expenditure (See Bond et al. (2021) for detailed explanations. A brief discussion on this is offered below). The accounting approach does not require any econometric assumption apart from that the marginal cost equals the average cost. This approach only needs cost and revenue data.

We do not observe prices and units of products sold, so only the production approach and the accounting approach are feasible. In fact, these are the methods used by many papers that infer firm-level markups using similar data as ours, such as De Loecker and Warzynski (2012), Liu (2019), Autor et al. (2020), De Loecker et al. (2020) and Baqaee and

Farhi (2020). We prefer using the accounting approach because we assume constant returns to scale and that the measurement errors are independent and identically distributed.

Dealing with the bias in the production approach is a lot of more difficult if not completely unfeasible. There are four sources of bias in the production approach under our setup when physical production and physical inputs are not observed and when firms have heterogeneous markups. The first one results from replacing production elasticities by revenue elasticities. If the revenue elasticities are consistently estimated, the estimated markups by the production approach should always be 1 (Bond et al. (2021)). Secondly, the assumption of variable input is very restrictive and it is almost impossible to find a truly variable input in data. Besides, the production approach also requires that the variable input do not affect demand and it can be common for inputs, such as labor inputs for marketing, to affect demand (Bond et al. (2021)). Most commonly used variable inputs are material and energy but we observe neither in our data. The last two sources are related to the consistency of estimated production elasticities using revenue data. In order to estimate production elasticities, the production approach needs to estimate production functions using Olley and Pakes (1996), Levinsohn and Petrin (2003), or Ackerberg et al. (2015). However, when revenue production is used in the place of physical production, Klette and Griliches (1996) demonstrates that heterogeneous markups can bias the estimated production elasticities downward. Last but no the least, even if one successfully corrects this bias by controlling for industry-level sales and prices, weak instruments can still plague the estimators (Bond et al. (2021)). Although Ridder et al. (2021) shows that estimated markups using revenue gives the correct dispersion but this requires using material as variable input. We only observe labor and capital. Since labor and capital are far from being variable, applying the production approach in our case is problematic.