Variable Markups and Misallocation in Chinese Manufacturing and Services*

Jinfeng Ge† Zheng (Michael) Song‡ Yangzhou Yuan§

December 13, 2016

Abstract

Cross-country comparison reveals an unusually small service sector in China. Using firm-level data from China's 2008 economic census, we find two facts that speak to a novel mechanism for misallocation within service and between manufacturing and service. First, compared with the manufacturing sector, there are more state-owned enterprises and fewer entrants in the service sector. Second, markups increase in firm size in both manufacturing and service, and the increase is more dramatic among service firms. We interpret these facts through the lens of a monopolistic competition model with heterogeneous firms and variable markups. Necessary and sufficient conditions are established for entry barriers and other frictions to cause misallocation via markups. We also extend the analysis to a multi-sector environment, where the model implies a new channel that translates asymmetric barriers to entry across sectors into sectoral markups differences, which, in turn, cause sectoral misallocation. To quantify the importance of the markups channel, the model is calibrated to match the observed firm size and markups distributions. The calibration finds big variations in entry barriers across industries. When reducing entry barriers for service firms to the extent observed for manufacturing firms, the model predicts a three-percentage-point increase in the service employment share.

*We thank John Hassler and seminar participants at the IIES, Department of Economics Stockholm University, Chinese University of Hong Kong, Fudan University for helpful comments.

†Fudan University, School of Economics, jinfeng_ge@fudan.edu.cn.

‡Chinese University of Hong Kong, Department of Economics, Email: zheng.michael.song@gmail.com.

§Stockholm University, Department of Economics, Email: yangzhou.yuan@ne.su.se.

1 Introduction

Less developed economies tend to have smaller service sector. China is no exception: Its service industries account for less than 40% of the total employment, about half of the share in the US. Yet, China's service sector deserves special attention for at least two reasons. First, China's service employment as a share of the total employment is among the smallest when compared to other countries at the same income level. When it comes to the service share in the non-agriculture sector, China is actually ranked at the lowest (see Figure 1). Second, the mirror image of the dwarfed service sector is China's extraordinarily large manufacturing sector, which has made the country "the world factory". Understanding why China has a unusually small service sector is among the first steps to rebalance the global economy.

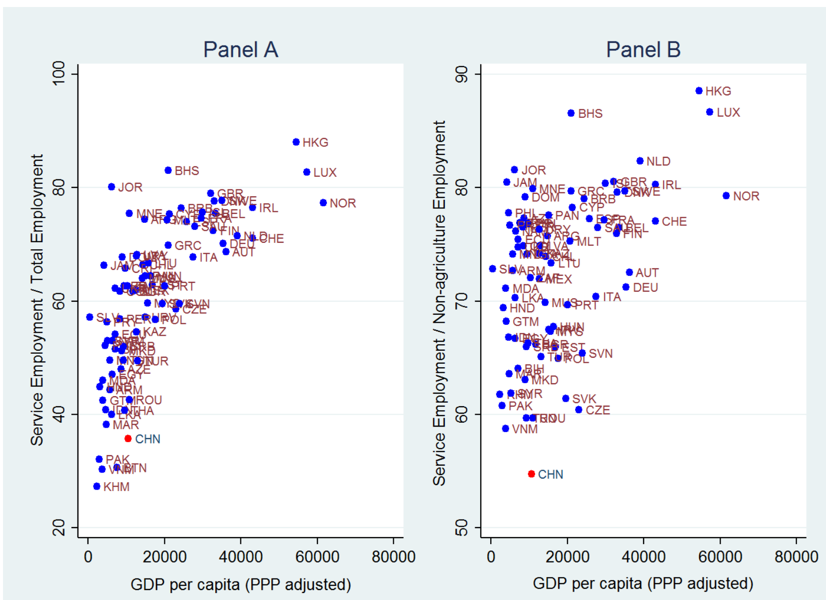

Figure 1: Panel A plots service employment as a percentage of total employment (y-axis) and GDP per capita (PPP adjusted, x-axis) in 2011. Panel B plots service employment as a percentage of non-agriculture employment (y-axis) and GDP per capita (PPP adjusted, x-axis) in 2011. Data source: World Bank.

Using firm-level data from China's 2008 economic census, we document two sets of facts that distinguish service from manufacturing and may speak to the underdeveloped service sector. First, the state share in the service sector is about three times as high as that in the manufacturing sector. Moreover, there is a robust negative correlation between the entry rate and the state share across both manufacturing and service industries. These patterns are consistent with the widely documented heavy regulations and high barriers to entry in many Chinese service industries (see, e.g., Rutkowski, 2015). We also find the dispersion of revenue-labor ratio among service firms to be 60 percent higher than that among manufacturing firms, suggesting more severe misallocation within service through the lens of the framework developed by Hsieh and Klenow (2009).[1]

Second, revenue-labor ratio increases in revenue among both manufacturing and service firms. The correlation is robust to within-industry heterogeneity in capital intensity. If we interpret variations in revenue-labor ratio, after controlling for capital intensity, as variations in markups, this finding would be in line with the recent models with variable markups (e.g., Zhelobodko et al., 2012; Dhingra and Morrow, 2014) that predict higher markups for larger firms. More importantly, we find the increase of revenue-labor ratio in revenue among service firms to be about 50 percent more than that among manufacturing firms. Since the average markups in an industry are mainly determined by markups changed by large firms, the steeper revene-labor ratio profile suggests higher average markups in service industries.

In summary, there are more SOEs and fewer entrants in China's service industries. Large service firms tend to charge higher relative markups compared to their counterparts in manufacturing. These facts point to a potential connection between markups and distortions caused by frictions such as barriers to entry. To lay down a theoretical foundation for such connection, we add entry barriers and other frictions into a standard monopolistic competition model with heterogeneous firms and variable demand elasticity. We establish necessary and sufficient conditions for entry barriers to affect the average markups and misallocation in an industry. Interestingly, if firm productivity follows Pareto distribution, entry barriers would have no effects due to the cancellation of the selection and variety effects. If, instead, firm productivity follows log normal distribution, the variety effect would dominate the selection effect, leading to the pro-competition effect that translates reduction in entry barriers or other frictions to lower average markups and alleviation of misallocation.

We apply the model to an multi-sector environment where the elasticity of substitution *across sectors* is constant. The novel finding is that under certain conditions, an asymmetry in frictions such as barriers to entry across sectors may lead to sectoral misallocation via markups. For instance, if the pro-competition effect exists, higher entry barriers in a sector would lead to higher average markups, resulting in too little resources allocated to the sector as opposed to those in the efficient allocation. Notice that in the standard models with constant markups,

[1]Hsieh and Klenow (2009) define revenue-labor ratio as labor revenue productivity. Revenue productivity is the product of physical productivity and a firm's output price. It should be equated across firms in the absence of distortions.

distorting the average productivity in a sector is the only channel through which entry barriers may cause sectoral misallocation. In contrast, our model entails both the standard productivity channel and the new markups channel. A more stark comparison is to consider the elasticity of substitution across sectors equal to one. The cancellation of the income and substitution effects would nullify the standard productivity channel. In other words, the markups channel would be the only channel through which entry barriers cause sectoral misallocation.

To examine the quantitative importance of the markups channel, we calibrate the model to match the observed firm size and markups distributions. The first finding is that although Pareto distribution fits well the right tail of the firm size distribution, it is a poor approximation for firms with size below the median level. Instead, log normal distribution that satisfies the necessary conditions for the markups channel fits well the observed distributions. When reducing entry barriers in service to the extent in manufacturing, the calibrated model predicts a three-percentage-point increase in the service employment share.

We find some auxiliary evidence for the mechanism that connects markups to entry barriers. China has experienced two major reforms in the late 1990s and early 2000s that greatly reduce entry barriers and foster domestic competition: (i) The restructuring of the state sector and (ii) the accession to WTO. Using the Annual Industrial Survey conducted by China's National Bureau of Statistics, we look for evidence whether such reforms led to reduction in markups among manufacturing firms. We find that the relative markups fell dramatically in most manufacturing industries from 1998 to 2007. Moreover, the reduction tends to be more dramatic in those industries with a higher share of newly established private firms, consistent with the mechanism that lowering barriers to entry would attract entrants and cut markups.

Our work contributes to the misallocation literature in two aspects. The theory part points out a new source of misallocation that arises from variable markups. When extended to the multi-sector environment, the theory illustrates the markups channel that translates asymmetry in frictions across sectors into sectoral misallocation. This provides a new perspective from which we understand resource allocation efficiency across sectors. In the quantitative part, we use firm-level data from both manufacturing and service to quantify the positive and normative implications of misallocation within an industry and across industries. Most of the existing work uses manufacturing firm data only and, hence, cannot address sectoral misallocation.

We are not the first on sectoral misallocation caused by markups differences. Epifani and Gancia (2011) show that trade liberalization may worsen sectoral misallocation by widening the gap of markups between tradable and nontradable sectors. In a broader sense, this paper is part of the literature on structural transformation (see, e.g., Herrendorf, Rogerson, and

Valentinyi, 2014). We provide a case study, illustrating how frictions can retard structural transformation.

When it comes to misallocation in China, the literature has well documented the policies favoring state-owned enterprises as an important source of misallocation.[2] There is also evidence linking entry barriers to the presence of state-owned enterprises across cities (see, e.g., Brandt, 2016). Bai, Hsieh and Song (2015) argue that large firms are more likely to be treated favorably through "special deals". Our findings suggest an asymmetry in favoring incumbent firms between manufacturing and service can distort sectoral allocation of resources.

Although Pareto distribution has widely been adopted in the literature, some recent studies explores some other distributions. For example, Fernandes et al. (2015) and Bas et al. (2015) argue for log normal distribution and Feenstra (2013) suggest bounded Pareto distribution. Our theory part offers another example of the gains from moving beyond Pareto. Also, we provide evidence for log normal distribution as well as its quantitative importance. This paper also contributes to a small but growing literature on frictions in the service sector. Song, Thomas, Wang and Xu (2016) develop a model where sales frictions can be reduced by accumulating retail capital. They calibrate the model to match some empirical moments from their survey on Chinese footwear firms. One of their main findings is that sales frictions can generate large markups heterogeneity. In that sense, sales frictions play a role similar to entry barriers in our paper, though the markups heterogeneity arises from an entirely different microfoudation.

Finally, our paper is related to the recent literature on variable markups. Following Zhelobodko, Kokovin, Parenti and Thisse (2012), Kichko, Kokovin and Zhelobodko (2013) and Dhingra and Morrow (2014), we adopt monopolistic competition model with additively separable utility functions that allow the demand elasticity to be increasing in quantity.[3] The contribution of our paper is two-fold. We characterize how frictions affect the average markups and explore its implications on sectoral misallocation. In addition, to deliver quantitatively sensible results, we examine all the four classes of utility in Dhingra and Morrow (2014): CARA, Expo, quadratic and HARA utility. Interestingly, only HARA utility has the capacity of fitting the main patterns of firm-level markups from China's firm-level data.

The rest of the paper is organized as follows. Section 2 documents a set of facts regarding misallocation and markups in manufacturing and service industries. We present the benchmark model in Section 3 and extend it to a multi-sector environment in Section 4. Section 5 calibrates

[2]See, e.g., Brandt and Zhu (2010), Song, Storesletten and Zilibotti (2011), Song and Wu (2014), and Hsieh and Song (2015).

[3]Alternatively, Edmond, Midrigan and Xu (2012) use a model of oligopoly and De Blas and Russ (2012) adopt Bertrand competition. See also Holmes, Hsu and Lee (2014) and Hsu, Lu and Wu (2016) for different ways of generating variable markups.

the model to the Chinese data and conducts counterfactual exercises. Further evidence is provided in Section 6. Section 7 concludes. All the proofs are in the online technical appendix.

2 The Facts

China's 2008 Economic Census covers five million firms.[4] We use two samples from the census, one for "above-scale" manufacturing firms and one for all service firms excluding financial institutions. The threshold for manufacturing firms is revenue of five million Yuan, the standard adopted by China's National Bureau of Statistical for the Annual Survey of Industrial Firms. The service sample has a lot more small firms, with more than two thirds below the scale. To make the two samples comparable, we drop all firms with revenue below five million Yuan. Table 1 reports the basic statistics of the two truncated samples. There are 29 and 34 two-digit manufacturing and service industries, and a total of 363 and 549 thousand above-scale manufacturing and service firms.

The literature has found huge disparity between state-owned and non-state-owned firms in Chinese manufacturing. We identify state-owned enterprise (SOEs henceforth) by either of the following two conditions: (i) the firm is registered as a state-owned unit; (ii) its state paid-in capital share is equal to or above 50 percent.[5] All non-state-owned enterprises are referred to private firms. The structural reforms initiated since the mid of the 1990s have greatly reduced the state presence in manufacturing (see, e.g., Hsieh and Song, 2015). Our census data shows that only 4 percent of manufacturing firms are state-owned. Their employment as a share of the total manufacturing employment is 13 percent (the top panel of Table 2). In contrast, we find a much larger state share in service, where 11 percent firms are state-owned and they account for 32 percent of the total service employment.

Entry barriers are obviously one of the reasons why the state sector remains sizable in some industries. Evidence that connects state shares to entry rates has been found in manufacturing (see, e.g., Brandt et al., 2012, Hsieh and Song, 2015 and Brandt et al., 2016). The census data allows us to look further for the evidence in both manufacturing and service industries.

[4] According to the regulations, economic census surveys "legal-person units, industrial units and self-employed individuals engaged in the secondary and tertiary industries within the territory of the People's Republic of China" (State Council, Decree No. 415, 2004). Legal-person units, including legal-person enterprises and government bodies, refer to the formally registered units that can independently bear civil liability. A legal-person enterprise may have single or multiple industrial units. 3.1 percent of the legal-person enterprises have multiple industrial units. In our sample, each observation is a legal-person enterprise, in which its industrial units are consolidated.

[5] Our definition of SOEs is broader than those using registration type only, for reasons stated in Hsieh and Song (2015). However, their definition is still broader than ours. They identify SOEs by those satisfying any of the two conditions or those with the state as the main shareholder. We cannot use their definition since the census data does not report shareholding information.

Specifically, we use the employment share of new private firms in each industry as a proxy for entry rate. New firms refer to the firms that were established after 1998 – i.e., those with age below 10 in 2008. Figure 2 shows that the employment share of new private firms (x-axis) in an industry is strongly correlated with the employment share of SOEs in that industry. The correlation is robust within manufacturing and service and between two sectors. Moreover, one can also see that service industries dominate manufacturing industries in the northwest area with low shares of new private firms but high shares of SOEs. All manufacturing industry but one (processing of petroleum, coking and processing of nuclear fuel) has the employment share of new private firms above 40 percent, while the share is below that level in 12 out of 34 service industries.

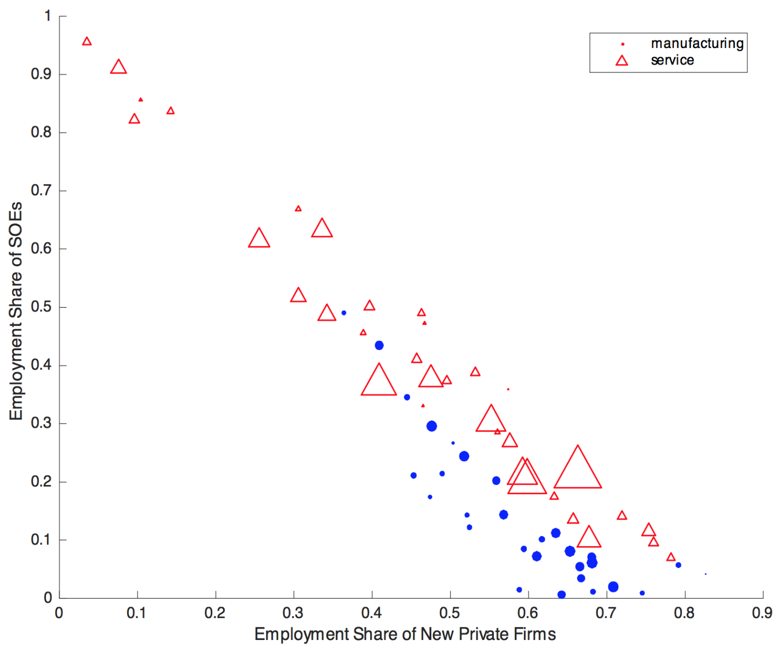

Figure 2: This figure plots the employment share of new private firms (x-axis) and the employment share of SOEs (y-axis) in manufacturing industries (dots) and service industries (triangles). The size of dots and triangles reflect the relative size of an industry by employment.

Figure 3 plots the distributions of revenue, revenue-labor ratio and revenue-capital ratio. We measure effective labor input by total wage bill. Revenue-labor ratio is thus equal to revenue per unit of total wages. Total assets are the only information related to capital in our samples. So, we use revenue per unit of total assets as a very crude measure for revenue-capital ratio. Most of the analysis in the following sections will be based on revenue and revenue-labor ratio. Revenue-capital ratio is only used for robustness check. Revenue, revenue-labor ratio and revenue-capital ratio are all in a relative sense, normalized by their corresponding median values in the industry. To reduce the influence of outliers, we drop observations with labor or revenue-capital ratio in the top or bottom 0.5 percentile in each industry.

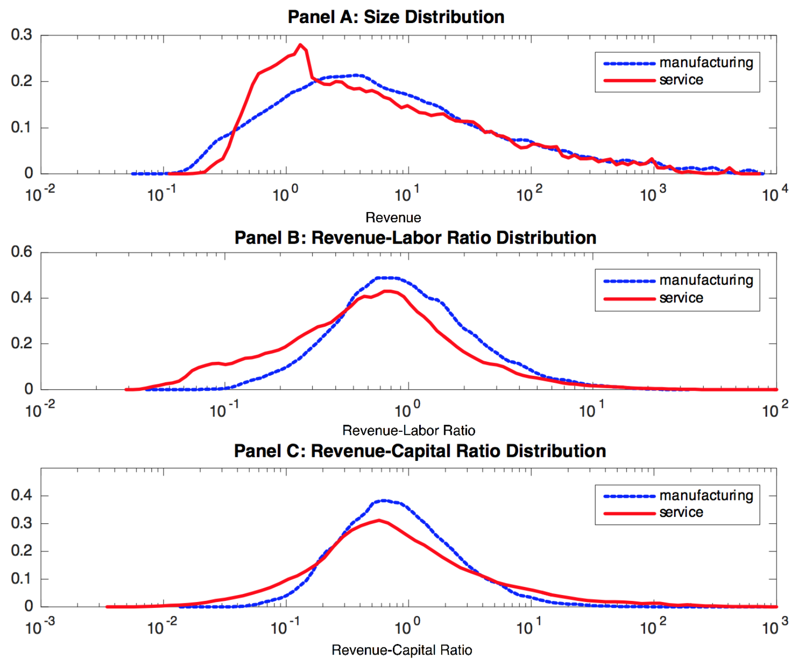

Figure 3: The dotted and solid lines in Panel A plot the revenue distribution of manufacturing and service firms, respectively. Revenue is normalized by the median value in each industry. Panel B and C plot the labor and revenue-capital ratio distributions. Revenue-labor ratio is revenue per unit of wage bill. Revenue-capital ratio is revenue per unit of assets. We normalize revenue, revenue-labor ratio and revenue-capital ratio by their median values in the industry. Observations are weighted by employment.

The revenue dispersions are similar between manufacturing and service firms. The variance of log revenue is 4.20 and 3.77 for manufacturing and service firms, respectively. The difference is about 10 percent. The difference of the revenue-labor ratio dispersion is much larger between the two sectors: The variance of log revenue-labor ratio among service firms is 1.18, about 60 percent higher than the variance of 0.74 for manufacturing firms. Bearing in mind the crudeness of the capital measure, we find an even larger difference in the revenue-capital ratio dispersion. The variance of log revenue-labor ratio is 1.20 and 2.75 for manufacturing and service firms, respectively. The misallocation literature would interpret these statistics as evidence for worse misallocation in the service sector.

We next group firms into percentiles by their revenue in each industry. The percentiles capture the within-industry firm size ranks. Panel A of Figure 4 plots the median of revenue-labor ratio in each percentile. It is immediate that revenue-labor ratio increases in revenue for both manufacturing and service firms. These findings are in line with the models where markups are variable and increase in firm size. More interestingly, the profile is steeper for service firms. The revenue-labor ratio of the top one percent manufacturing firms relative to that of the bottom one percent manufacturing firms is 3.5, while the ratio is 4.7 for service firms. Column 1 of Table 2 regresses log revenue-labor ratio against log revenue and the interaction term between log revenue and the dummy variable for service firms. Both of the estimated coefficients are positive and highly significant. In particular, the results suggest the revenue-labor ratio profile for service firms is about 50 percent steeper than that for manufacturing firms.

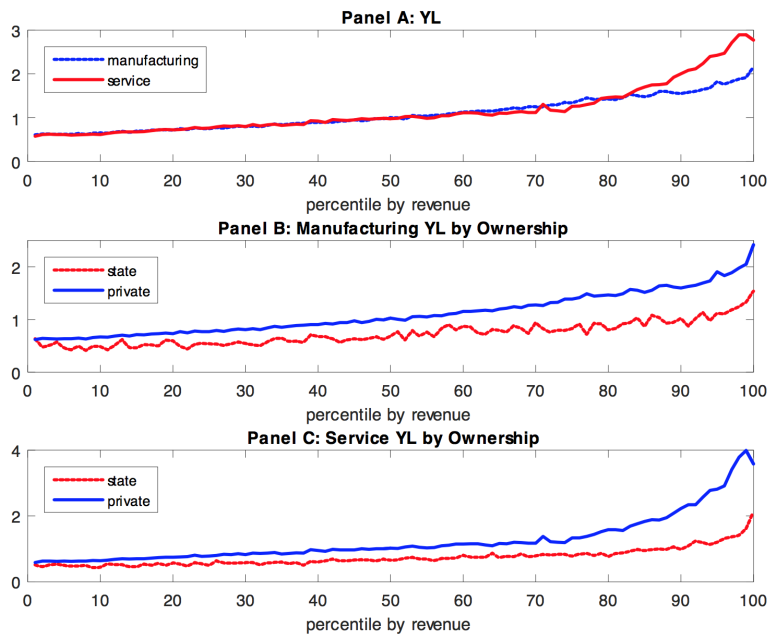

Figure 4: We group firms into percentiles by their revenue in an industry. Panel A plots the median revenue-labor ratio in each percentile for manufacturing firms (dotted line) and service firms (solid line). Panel B plots the median revenue-labor ratio in each percentile for state-owned manufacturing firms (dotted line) and private manufacturing firms (solid line). Panel C plots the revenue-labor ratio profiles for service firms.

A different reading of Panel A is to think of the upward-sloping revenue-labor ratio profile as larger firms adopting less labor-intensive technology. The hypothesis would predict a negative correlation between revenue-capital ratio and firm size.[6] This is obviously inconsistent with the findings in Panel A of Figure 5. As a robustness check, we add the capital-labor ratio into

[6]This can be seen through the lens of (5) below under $\alpha' < 0$.

the benchmark regression. One can see from Column 2 of Table 2 that the capital-labor ratio has the expected sign. Its effect on revenue-labor ratio is stronger for manufacturing firms. The estimated coefficients of our main interests appear to be very robust.

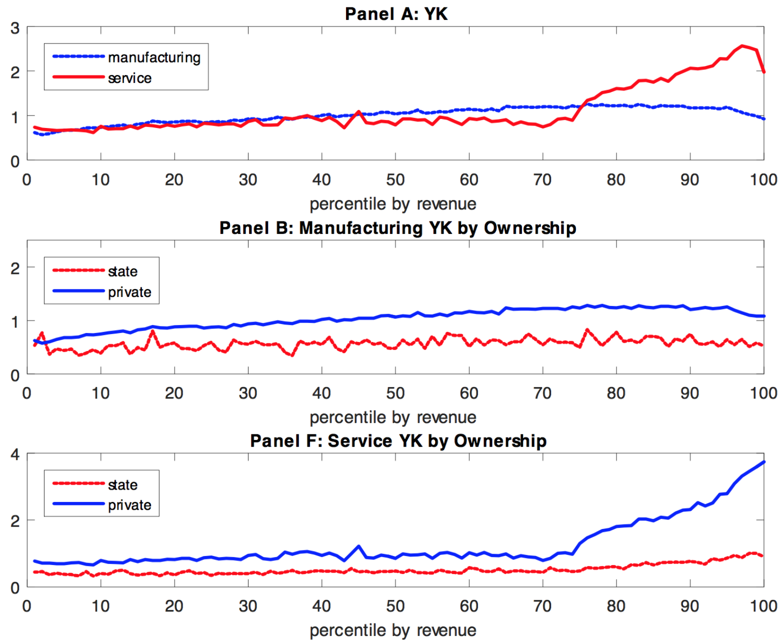

Figure 5: We group firms into percentiles by their revenue in an industry. Panel A plots the median revenue-capital ratio in each percentile for manufacturing firms (dotted line) and service firms (solid line). Panel B plots the revenue-capital ratio profiles for manufacturing firms. Panel C plots the revenue-capital ratio profiles for service firms.

Panel B in Figures 4 and 5 plots the revenue-labor ratio and revenue-capital ratio profiles by ownership for manufacturing firms. Consistent with the findings in the literature, private firms are associated with higher revenue-labor ratio and revenue-capital ratio. Panel C plot the results for service firms. There, revenue-labor ratio and revenue-capital ratio are also higher for private firms. The revenue-labor ratio profile is always upward-sloping, regardless of sector or ownership. Columns 3 and 4 of Table 2 confirms the finding by including the SOE dummy and its interactions with log revenue and the capital-labor ratio as additional controls. The revenue-capital ratio profile shows a similar pattern, though it becomes much weaker for state-owned manufacturing firms.

To conclude, we find the following facts:

1. The state employment share is a lot higher in service industries than that in manufacturing industries. There is a negative correlation between the state share and entry rate across both manufacturing and service industries.

2. The firm size distribution is similar between the manufacturing and service sectors, while the revenue-labor ratio and revenue-labor ratio distributions are more dispersed in service.

3. The revenue-labor ratio is increasing in revenue, regardless of sector and ownership. The revenue-labor ratio profile is substantially steeper for service firms than that for manufacturing firms.

3 The One-Sector Model

In this section, we lay out a simple monopolistic competition model with heterogeneous firms and variable markups. Our main purpose is two-fold. First, the one-sector version of the model demonstrates a novel channel through which frictions may lead to higher markups. A set of necessary and sufficient conditions will be established for the channel to function. The model will be extended to a multi-sector environment, where different magnitudes of frictions across sectors may lead to sectoral misallocation. Second, using Chinese census data, we estimate the model and conduct counterfactual experiments to quantify the importance of the markup channel. In addition, the model generalizes the framework in Hsieh and Klenow (2009) by introducing variable markups. Our approach can thus isolate misallocation caused by variable markups from the Hsieh-Klenow estimator of misallocation.

There are N firms. Each produces a single variety. Labor is the only input factor, an assumption that will be relaxed below. The production technology is $l_i = c_i l\left(q_i\right)$, where $l' > 0$, $l'' \geq 0$, q_i is the quantity of the good, l_i denotes the labor input and $1/c_i$ captures the firm TFP. We abstract capital input for simplicity in the benchmark model and will bring it back for robustness check.

Consider a representative consumer who has the following preferences over differentiated goods:

$$U = \int_0^N u\left(q_i\right) di, \tag{1}$$

where u is continuously differentiable, monotonically increasing , strictly concave and $u\left(0\right) = 0$. Here we do not assume specific form of $u(\cdot)$.[7]

The downward-sloping demand curve is characterized by $p\left(q_i\right) = u'\left(q_i\right)/\lambda$, where λ is the shadow price of consumer constraint. [8] And demand elasticity is:

$$-\frac{\mathrm{d}q_i}{\mathrm{d}p_i}\frac{p_i}{q_i} = -\frac{u'\left(q_i\right)}{q_i u''\left(q_i\right)}$$

[7] The CES utility is a special case with an iso-elastic $u(\cdot)$.

[8] See Appendix 8.2 for detailed derivation of household problem.

We define the inverse of the demand elasticity as $\mu(q_i)$:

$$\mu(q_i) \equiv -\frac{q_i u''(q_i)}{u'(q_i)}. \tag{2}$$

Under the general assumptions of $u(\cdot)$, $\mu(q_i)$ is allowed to be a variable function of q_i. As a consequence, firms charge variable markups under variable demand elasticity. The special case is CES utility, which implies a constant μ and, hence, constant markups. [9] The variable demand elasticity here establishes the microfoundation that links markups to firm size. In a symmetric case where q_i is the same for all i, $\mu(q_i)$ represents the inverse of the elasticity of substitution. μ is also referred to as the relative love for variety in Zhelobodko *et al* (2012) or private markup in Dhingra and Morrow (2014).[10]

All firms face the same demand curve, $p_i = p(q_i)$, and need to pay taxes. The after-tax revenue is $(1-\tau_i) p_i q_i$. τ_i has two components: (i) the rate that applies to all firms; and (ii) the rate that is firm-specific. Without loss of generality, the common tax rate is normalized to zero. We then interpret τ_i as frictions caused by distortionary policies. Accordingly, we will refer to misallocation caused by τ_i as policy-induced misallocation.[11]

Firms make production decision by

$$\max_{q_i} (1-\tau_i) p(q_i) q_i - w c_i l(q_i). \tag{3}$$

Define markup as price over marginal cost, the first-order condition implies

$$\underbrace{\frac{p(q_i)}{w c_i l'(q_i)}}_{\text{Markups}} = \frac{1}{1-\tau_i} \frac{1}{1+\frac{p'(q_i) q_i}{p(q_i)}} = \frac{1}{1-\tau_i} \frac{1}{1-\mu(q_i)}, \tag{4}$$

(4) shows that markups, the LHS of the equation, are co-determined by τ_i and $\mu(q_i)$.

(4) can be rewritten as

$$\log \underbrace{\frac{p(q_i) q_i}{w c_i l_i}}_{\text{revenue-labor ratio}} = -\log(1-\tau_i) - \log((1-\mu(q_i)) e(q_i)), \tag{5}$$

where $e(q_i) \equiv l(q_i) / (l'(q_i) q_i)$ is the output elasticity. When the demand and output elasticities are invariable, (5) reduces to the one in Hsieh and Klenow (2009). Another way to think of (5) is to interpret the LHS as the inverse of labor income share. By adding τ_i, (5)

[9] With CES power utility: $u(q) = q^{\frac{s-1}{s}}$, we have $\mu(q_i) = 1/s$.

[10] Under Assumption 1 stated below, consumers will perceive varieties as being less differentiated when their consumption is higher.

[11] Alternatively, one may interpret τ_i as firm-specific labor income tax rate, which is observationally equivalent to τ_i in the model. For expositional ease, we assume a common wage rate, denoted by w, for all firms and load all the distortions to τ_i.

extends the well-known formula that equates markups to the ratio of output elasticity to factor share (see, e.g., De Loecker and Warzynski, 2012). Finally, variable markups would invalidate the welfare calculation in Hsieh and Klenow (2009) since τ_i would distort resource allocation through variable markups, a channel that is absent in their model.

For simplicity, we specifically assume that $l(q_i) = q_i$ through the following sections. Since $p(q_i)$ is a function of $u'(q_i)$, rewrite (4) and $u'(q_i)$ can be expressed as a function of $\frac{c_i}{1-\mu(q_i)}$. We make the following assumption throughout the rest of the paper.

Assumption 1: $0 \leq \mu(\cdot) \leq 1$, and $\mu'(\cdot) > 0$, where μ is defined in (2).

Lemma 1 *Under Assumption 1, q_i, markups, revenue and profits are all strictly decreasing in c. revenue-labor ratio is hence increasing in revenue.*

See Appendix 8.1 for proof. Lemma 1 is consistent with the third fact in Section 2.

3.1 Entry Barriers

We now consider a particular type of frictions: barriers to entry. Active firms come from a pool of $\bar{N}$ potential producers. We use $\bar{N}$, as a measure of entry barriers. [12] To begin with, we shut down policy-induced misallocation (i.e., $\tau_i = 0\ \forall i$). Each potential producer from $\bar{N}$ draws a productivity c from a distribution $G(c)$.

For simplicity, we assume there is no fixed costs for production. The firms that are indifferent between entering and staying outside the market must make zero profits. Without fixed cost, existence of these break even firms requires $u'(0) < +\infty$.[13] Notice CES utility does not satisfy this condition and we are focusing on utility functions with variable elasticity.[14] Moreover, there exists a unique cutoff productivity, denoted by c_D, such that only firms with $c \leq c_D$ will make profits and be active. A formal proof is provided in the online appendix. The zero-profit condition implies that q_i associated with the cutoff productivity is also zero.

The first-order condition becomes:

$$u'(q_i)(1 - \mu(q_i)) = c_i \lambda, \tag{6}$$

[12] $\bar{N}$ is similar to fixed entry cost setting. With entry cost setting, potential entrepreneurs pay entry cost, draw productivity and then start production. With free entry condition, entry cost equals to expected profits after entry. In equilibrium, a higher entry cost indicates a higher expected profits after entry. As a consequence, less firms can survive with higher expected profits in equilibrium. Chaney (2008) uses similar setting. Instead of free entry condition, total mass of potential entrants in country is proportional to country size.

[13] $u'(0) < +\infty$ implies demand curve intersects vertical axis when $q = 0$. At intersection points, firms face zero demand and make zero profits. However, when $u'(0) = +\infty$, all firms are active.

[14] All utilities listed in section 5 satisfy this condition.

which implies

$$\frac{u'(q_i)(1-\mu(q_i))}{u'(0)(1-\mu(0))} = \frac{c_i}{c_D}., \tag{7}$$

We can establish that $q_i = q(c_i/c_D)$ and $q' < 0$ (see the online appendix for proof). The lowest markups are thus equal to $1/(1-\mu(0))$.

There exists a competitive labor market where labor supply is inelastic and equal to L. The market clearing condition pins down c_D:

$$\bar{N}\int_0^{c_D} cq\left(\frac{c}{c_D}\right) dG(c) = L. \tag{8}$$

Define Φ as the average markups.

$$\begin{aligned}\Phi \equiv \frac{Y}{L} &\equiv \frac{\int_0^{c_D} \frac{c}{1-\mu(q(c/c_D))} q\left(\frac{c}{c_D}\right) dG(c)}{\int_0^{c_D} cq\left(\frac{c}{c_D}\right) dG(c)} \qquad (9)\\ &= \int_0^{c_D} \underbrace{\frac{cq\left(\frac{c}{c_D}\right)}{\int_0^{c_D} cq\left(\frac{c}{c_D}\right) dG(c)}}_{\text{employment share}} \underbrace{\frac{1}{\left(1-\mu\left(q\left(\frac{c}{c_D}\right)\right)\right)}}_{\text{markups}} dG(c).\end{aligned}$$

For analytical convenience, we define $\Psi(c/c_D)$ as the relative productivity distribution of active firms, where $c/c_D \in [0,1]$.

Proposition 1 *Consider two distributions, Ψ^1 and Ψ^2. If Ψ^1 first-order stochastically dominates Ψ^2,*

(i) the average markup associated with Ψ^2 will be higher than that associated with Ψ^1;

(ii) the revenue-labor ratio distribution associated with Ψ^2 will be more dispersed than that associated with Ψ^1. [15]

See the online appendix for proof. The first-order stochastic dominance implies that for any $x \in [0,1]$, there will be more active firms – i.e., those with the relative productivity $c/c_D > x$, in the equilibrium with Ψ^2. In other words, the equilibrium with Ψ^2 has a higher employment share of high-productivity firms. Since high-productivity firms are larger and charge higher markups, the composition effect leads to higher average markups. For example, assume that c follows a Pareto distribution with $G(c) = (c/\bar{c})^{\kappa}$, where $c \in [0,\bar{c}]$. Then, $\Psi(c/c_D) = (c/c_D)^{\kappa}$. The first-order stochastic dominance holds if $\kappa_1 > \kappa_2$. Proposition 1 guarantees that a lower κ will lead to higher average markups and more dispersed revenue-labor ratio.

Assumption 2: $\Psi\left(c/c_D^1\right)$ first-order stochastically dominates $\Psi\left(c/c_D^1\right)$ if $c_D^1 < c_D^2$.

[15] We use the concept Lorenz Domination to measure dispersion, see Appendix 8.2 for detailed definition.

Proposition 2 *Under Assumption 2, a high $\bar{N}$ (i.e., low entry barrier) or a less dispersed τ_i will reduce the average markup and the dispersion of revenue-labor ratio.*

Proposition 2 is a key that connects frictions to markups. The selection effect increases the productivity and output of active firms, resulting in higher average markups. However, there is an opposite effect via variety. A higher $\bar{N}$ leads to more varieties, implying less labor allocated to each variety, which, in turn, lowers markups. A counterexample to Assumption 2 is Pareto distribution. We can show that if c is Pareto distributed, a higher $\bar{N}$ will reduce c_D but leave $\Psi\left(c/c_D^1\right)$ unchanged (see the online appendix for proof). In this case, the selection effect and the variety effect cancel out with each other under the Pareto distribution.

Melitz and Ottaviano (2008) and Behrens et al. (2014) establish that under quasi-linear or CARA preferences and Pareto distribution, the average markup is independent of trade frictions. Our result is more general since it applies for all additive preferences. Moreover, it suggests that the fat-tailed nature of the Pareto distribution is the key: It strengthens the selection effect. If c instead follows a log-normal distribution with a less fat tail that satisfies Assumption 2, the selection effect will not be strong enough to balance the variety effect.

4 The Multi-Sector Model

We apply the above model to a multi-sector environment. Assume there are J sectors and U is a CES aggregator of U_j of sector j:

$$U = \left[\sum_{j}^{J} \gamma^j \left(U^j\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}},$$

where $\sum_j^J \gamma^j = 1$, $U^j = \int_0^1 u^j\left(q_i^j\right) di$ and $\sigma > 0$ is the elasticity of substitution across sectors. There are $\bar{N}^j$ potential firms in sector j, where the productivity distribution is $G^j(c)$. The cutoff-level productivity in sector j is c_D^j. The labor market clearing condition (8) becomes

$$\sum_j \bar{N}^j \int_0^{c_D^j} cq\left(\frac{c}{c_D^j}\right) dG^j(c) = \sum_j L^j = L. \tag{10}$$

Our focus is on the sectoral allocation of resources between manufacturing and service. To sharpen the results, we establish the following proposition in a two-sector environment with $J = 2$. Sector 1 and 2 are referred to as manufacturing and service, respectively. All the results can easily be extended to $J > 2$. For analytical convenience, we will also assume that u^j and G^j are the same across sectors. The online appendix proves that Proposition 1 and 2 carry over to the multi-sector environment.

Proposition 3 *Under Assumption 2 and $\sigma \geq 1$, a higher $\bar{N}^1$ (lower entry barrier in manufacturing) or a less dispersed τ_i^1 will increase the average markup in service and cause labor to flow from service to manufacturing.*

Under Assumption 2, a lower entry barrier in manufacturing will increase the average productivity and reduce the average markup in the sector (Proposition 2). Holding the relative price of manufacturing goods unchanged, the lower average markup would increase the demand for manufacturing goods, causing labor to flow towards the manufacturing sector. We refer to this as the markups channel, which is absent in the models with constant markups. In addition, the standard productivity channel is also present. The higher average productivity in manufacturing lowers the relative price of manufacturing goods. When $\sigma > 1$, the substitution effect of the lower relative prices will dominate the income effect. This will also cause labor to reallocate from service to manufacturing. In the following quantitative exercise, we will assume $\sigma = 1$ in the benchmark case so that the income and substitution effects cancel out each other. The overall effect becomes ambiguous when $\sigma < 1$.

To see the mechanism more transparently, we obtain the following equation that governs sectoral labor allocation:

$$L^j = \Gamma \left[\frac{\left(\gamma^j\right)^\sigma \left(c_D^j\right)^{1-\sigma}}{\Omega^j} \right]$$

where Ω^j is strictly increasing in the average markups (see the online appendix for details). The term with c_D^j captures the standard productivity channel, where the average productivity in sector j increases in the cutoff productivity, $1/c_D^j$. The direction of the productivity effect hinges on the value of σ. The term with Ω^j captures the markups channel. In the special case with $\sigma = 1$, the markups channel would be the only mechanism through which entry barriers affect resource allocation across sectors.

It is worth mentioning what would happen if the productivity distribution is Pareto. There, a lower entry barrier or a less dispersed output wedge in manufacturing will still increase the average productivity (i.e., $1/c_D^j$) in manufacturing. However, Ω^j that reflects the average markup remains unchanged. The comparative statics is analogous to that in Ngai and Pissaridis (2007). In other words, Pareto productivity distribution will shut down the markup effect and, thus, the new channel for sectoral resource allocation with variable markups.

5 Quantitative Results

5.1 Utility Function

We first examine the empirical predictions of the four classes of utility function in Dhingra and Morrow (2014) that satisfy the assumptions that $\mu(q) \in [0,1]$ and $\mu'(q) > 0$. Although all the utility functions can generate a upward-sloping size-markup profile, their predictions on the dispersions of size and markup turn out to be very different.

- CARA Utility:
$$u(q) = 1 - \exp(-\alpha q),$$
with $\alpha > 0$ and $\mu(q) = \alpha q$. In the limiting case with $c \to 0$, $q \to 1/\alpha$ as $\mu(q) \in [0,1]$ and, hence, $p/c \to \infty$. In words, the most productive firm will charge infinitely high markups. This is obviously inconsistent with the empirical regularity that labor income share is quantitatively sizable even among the top one percent firms in each industry. Moreover, firm employment is not monotonically increasing in firm productivity. As $c \to 0$, $l = cq \to 0$.

- Expo Utility:
$$u(q) = 1 - \exp\left(-\alpha q^{1-\rho}\right),$$
with $\alpha > 0$, $\rho \in (0,1)$ and $\mu(q) = \alpha(1-\rho)q^{1-\rho} + \rho$. The assumption that $\mu(q) \in [0,1]$ implies $q \in \left[0, 1/\alpha^{1-\rho}\right]$ and $c \to 0$, $p/c \to \infty$. So, if ρ is sufficiently large, Expo utility can generate large output and revenue dispersion. Bùt a larger ρ also implies higher markups charged by the firms at the cutoff-level productivity: $p/c = 1/(1-\rho)$. Therefore, to match the observed size dispersion, Expo utility has to resort to large ρ that will generate unrealistically high markups for the smallest firms. Similar to CARA utility, the relationship between firm productivity and employment is ambiguous.

- Quadratic Utility:
$$u(q) = \alpha q - \frac{\beta}{2}q^2,$$
where $\alpha > 0$, $\beta > 0$ and $\mu(q) = \beta q/(\alpha - \beta q)$. The assumption that $\mu(q) \in [0,1]$ implies $q \in [\beta/\alpha, \beta/(2\alpha)]$. In the limiting case with $c \to 0$, $q \to \beta/(2\alpha)$ and, hence, $p/c \to \infty$. So, the size-markups profile implied by quadratic utility is similar to that implied by CARA utility and, hence, subject to the same critique.

- HARA Utility:
$$u(q) = \frac{[q/(1-\rho) + \alpha]^{\rho} - \alpha^{\rho}}{\rho/(1-\rho)}, \tag{11}$$

where $\alpha > 0$, $\rho \in (0,1)$ and $\mu(q) = q/(\alpha + q/(1-\rho))$. In the limiting case with $c \to 0$, q and $pq \to \infty$ but $p/c \to 1/\rho$. Moreover, $q = 0$ and $p/c = 1$ at the cutoff productivity. These are two important properties. First, HARA utility can generate large output dispersion. Second, revenue and employment are monotonically increasing in productivity.

It is also worth noting that ρ has two opposite effects on the average markups in a sector. On the one hand, a lower ρ tends to increase the average markups by making larger firms charge higher markups. On the other hand, it also implies less substitutability across varieties, making the firm size distribution less dispersed. This tends to lower the average markups. We show in the online appendix that the former effect dominates the latter effect and the average markups are always decreasing in ρ.

We then adopt HARA utility in our quantitatively exercise and assume ρ^j to be industry-specific. The model predicts a upperbound of $1/\rho^j - 1$ for markups.[16] We will calibrate ρ^j to match the observed markups and revenue distributions. In addition to ρ^j, HARA utility has another parameter α^j. But it does not affect any of the revenue, employment and markup distributions.[17] So, we simply set it to unity. The results are robust to various values of α^j.

5.2 Calibrating Productivity Distributions

When markups are constant, the firm size distribution would be isomorphic to the underlying productivity distribution. Our simulations show that although variable markups affect firm size, the difference between the firm size and productivity distributions are quantitatively small. In particular, if productivity follows Pareto or log normal, then the firm size distribution would also be nicely fitted by a Pareto and log normal distribution. The similarity allows us to check if the productivity distribution satisfies Assumption 2 by looking at the firm size distribution.

Although Pareto distribution has been widely adopted in the literature, some recent studies challenge the applicability of the assumption.[18] Figure 6 plots the log size and log rank for manufacturing and service firms, where size is employment relative to the median value in the industry. It is immediate that the power law doesn't apply for firms with employment below the industry median. Following Eeckout (2004), we use the Kolmogorov-Smirnov (K-S) test to check the goodness of fit for log normal distribution. This gives the K-S statistic of 0.031

[16] This can easily be extended to the Cobb-Douglas production technology with capital and intermediate inputs. To see this, denote $e < 1$ the labor output elasticity. The corresponding labor income share for the firms with $c \to 0$ and those with the cutoff productivity is ρe and e, respectively.

[17] We can write q^j/α^j as a function of c^j/c^j_D and ρ^j.

[18] See Feenstra (2013), Fernandes et al. (2015) and Bas et al. (2015).

and 0.054 for manufacturing and service firms, respectively. The corresponding p value is less than 1 percent in both cases. That is to say, log normal distribution fits well the firm size distribution in the census.

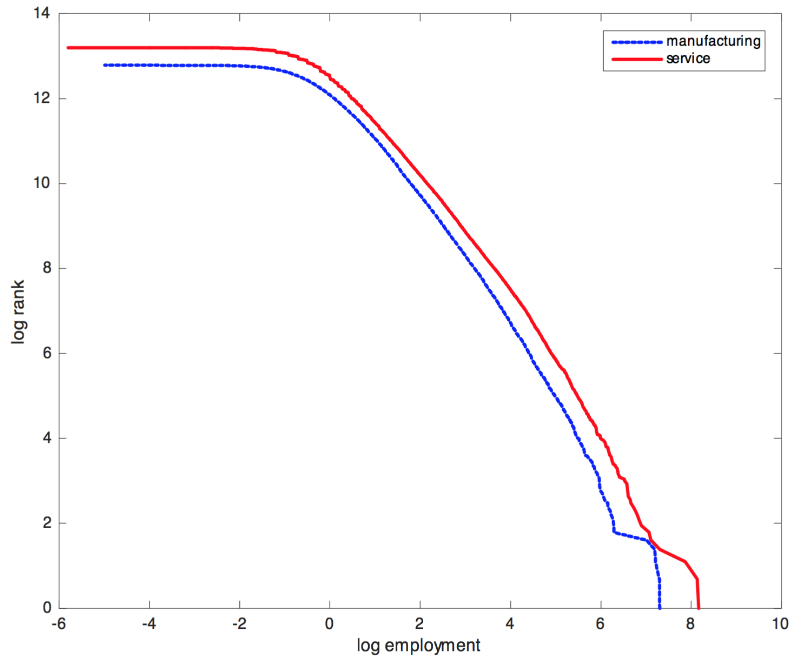

Figure 6: This figure plots log employment (relative to the median employment in the industry) against log rank. The dotted and solid lines are for manufacturing and service firms, respectively.

Let $G^j(c)$ follow log normal distribution, where

$$G^j(c) = \frac{1}{2} - \frac{1}{2}\operatorname{erf}\left(\frac{-\ln c - \varphi^j}{s^j\sqrt{2}}\right),$$

where φ^j and s^j denotes the mean and standard deviation, respectively. $\exp\left(\varphi^j\right)$ is normalized to unity. Then,

$$\Psi^j\left(x^j\right) = \frac{1 - \operatorname{erf}\left(1/\left(s^j\sqrt{2}\right)\left(-\ln x^j - \ln c_D^j\right)\right)}{1 - \operatorname{erf}\left(1/\left(s^j\sqrt{2}\right)\left(-\ln c_D^j\right)\right)},$$

where $x^j = c/c_D^j$. We then estimate c_D^j and $s^j\sqrt{2}$by matching the revenue and markup distributions. Specifically, we stimulate 10,000 firms for each industry and group firms into percentiles by their revenue. We then generate the revenue and markups profiles, which plot the median value of revenue or markups in each percentile relative to that in the bottom percentile. Nonlinear least square is used to estimate c_D^j and $s^j\sqrt{2}$ to minimize the distance between the stimulated and empirical profiles.

5.3 Calibrating Entry Barriers

We assume that the number of potential firms, $\bar{N}^j$, equals $\bar{N}^{S,j} + \bar{N}^{P,j}$, where the superscript S or P represents state-owned and private firms, respectively. Without loss of generality, we let $N^{S,j} = \bar{N}^{S,j}$, namely, there is no entry barrier for SOEs and all potential SOEs are active. Private firms face entry barriers. $N^{P,j}/\bar{N}^{P,j}$ measures the magnitude of entry barriers in industry j. To back out $\bar{N}^{P,j}$, we rewrite the labor market clearing condition as

$$\bar{N}^j = \bar{N}^{S,j} + \bar{N}^{P,j} = \frac{L^j}{1/2 \cdot c_D^j \left[1 - 1\,\mathrm{erf}\left(\frac{-\ln c_D^j}{s^j\sqrt{2}}\right)\right] \int_0^1 x q^j(x)\, d\Psi^j(x)}.$$

Given c_D^j and s^j, the above equation would pin down $\bar{N}^j$ and, hence, $\bar{N}^{P,j}$.

5.4 Calibrating Preference Parameters

We set $\sigma = 1$ as the benchmark case. The household consumption decision implies c_D^j is a function of $\gamma^j \left(U^j\right)^{-1/\sigma}$:

$$c_D^j = \Upsilon \left[\gamma^j \left(\bar{N}^j G^j\left(c_D^j\right) \int_0^1 u^j\left(q^j(x)\right) d\Psi^j(x)\right)^{-1/\sigma}\right]$$

Given c_D^j, s^j and $\bar{N}^j$, we can back out γ_i by the above equation and the constraint that $\Sigma_j \gamma^j = 1$.

5.5 Results

The results are reported in Table 3. The fitness is very good. The average R square is 0.94 and 0.95 for the markups and sales profiles, respectively. The calibrated economy implies severe sectoral misallocation and large efficiency losses. We now conduct counterfactual experiments to illustrate the quantitative importance of one particular distortion over the extensive margin. 4.1% and 25.4% of manufacturing and service firms are state-owned. In the counterfactual exercise, we increase the number of potential private firms in each service industry by the same proportion such that the number of active SOEs relative to that of active private firms in the service sector is identical to that ratio in the manufacturing sector. The results can be seen from Figure 7. The first finding is that the extensive margin can generate quantitatively sizable effects. The average markups in the real estate industry, for instance, would drop by 9.2 percent. Figure 7 also reveals a linear pattern between the percentage changes in the average markups and total employment. The real estate industry would gain employment by 17%. Overall, the employment share of the service sector would increase by 3 percentage points. The aggregate welfare gain would be increased by 32 percent. The big increase is largely driven by the increase in the average productivity. The decomposition shows that our markups channel accounts for about one fifth of the welfare gain. In other words, holding the average productivity constant, reducing entry barriers would increase the aggregate welfare by 6 percent through the variable markups channel.

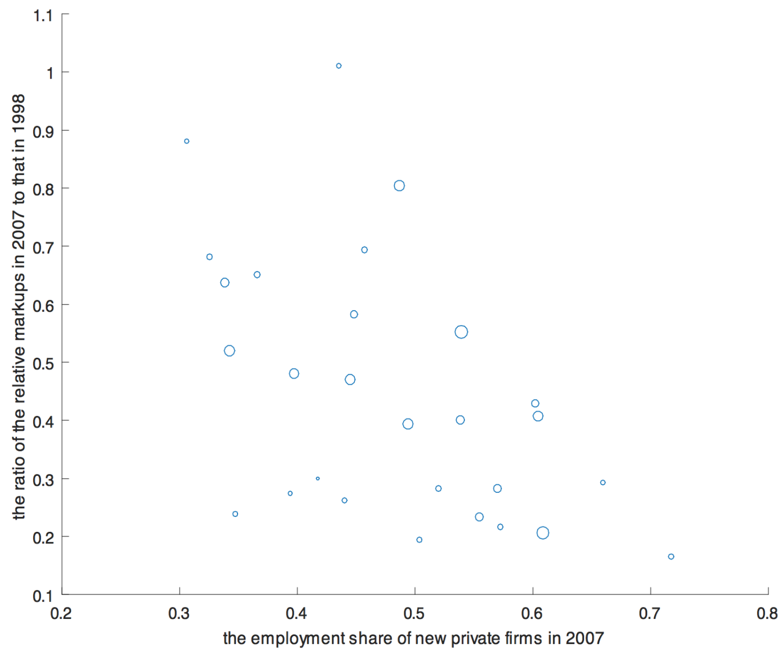

Figure 7: This figure plots the results of the counterfactual experiment (see the text for details). The x- and y-axis represent the percentage change of the average markups and total employment in each service industry.

6 Discussion

We discusses in this section some of the potential reasons why entry barriers tend to be lower in Chinese manufacturing industries. Moreover, we will provide evidence that connects entry barriers to changes in the average markups. The Chinese authorities have been implementing the SOE reforms under the policy slogan of "Grasping the Large, Letting Go of the Small" since the mid of the 1990s. While many SOEs in manufacturing industries were shut down or privatized (Hsieh and Song, 2015), some service industries such as banking, telecommunication and transportation are still heavily regulated to favor incumbent SOEs. The official reason is to support the industries of "vital importance to the economy and people's livelihood." China's accession to WTO in 2001 is another major breakthrough that lowers barriers for Chinese firms to enter the global markets and enhances domestic competition (see, e.g., Brandt et al., 2014). While most trade barriers have been removed for many manufacturing industries, it is often hard to open up service industries to foreign competition for political reasons. This has been seen in many countries.[19]

These observations prompt us to look into the correlation between entry barriers and changes in the average markups across industries. The model predicts that (i) the average markups should decline in manufacturing industries due to the SOE reforms and accession to WTO; (ii) the average markups should decline more in the industries where the share of new firms is higher. We use the firm-level data from the annual industrial survey conducted by China's National Bureau of Statistics from 1998-2007. The annual industrial survey allows us to keep track of changes in the average markups. The disadvantage is that the survey covers industrial firms only.

We do not have a direct measure on the average markups. Under HARA preferences, the firms with the cutoff productivity would always make zero profits. Higher barriers to entry increase the average markups by increasing the markups charged by high-productivity firms. Hence, we proxy the average markups by the relative markups – i.e., the ratio of revenue-labor ratio of the firms with revenue in the top five percentiles to that of the firms with revenue in the bottom five percentiles. Figure 8 plots the employment share of new private firms (x-axis) and the ratio of the relative markups in 2007 to that in 1998 (y-axis) across manufacturing industries. First notice that the relative markups fall dramatically in most manufacturing industries in the sample period from 1998 to 2007. There is only one industry (manufacture of rubber) where the relative markups go up . Moreover, the negative correlation illustrated in the figure attests to the mechanism that lower barriers to entry may reduce the average markups.

7 Conclusion

To recapitulate, this paper contributes to the literature in two aspects. On the theoretical front, we establish the conditions for distortions over extensive and intensive margins to affect the average markups. Moreover, we illustrate how the within-industry distortions can lead to sectoral misallocation. On the other hand, we show the quantitative importance of the channel by calibrating the model to the Chinese economy. In particular, we find that removing entry barriers for private service firms to the extent for private manufacturing firms would be able to increase the employment share of China's service sector by three percentage points. Such

[19]One example is the restrictions on foreign investment in services (see, e.g., Rutkowski, 2015).

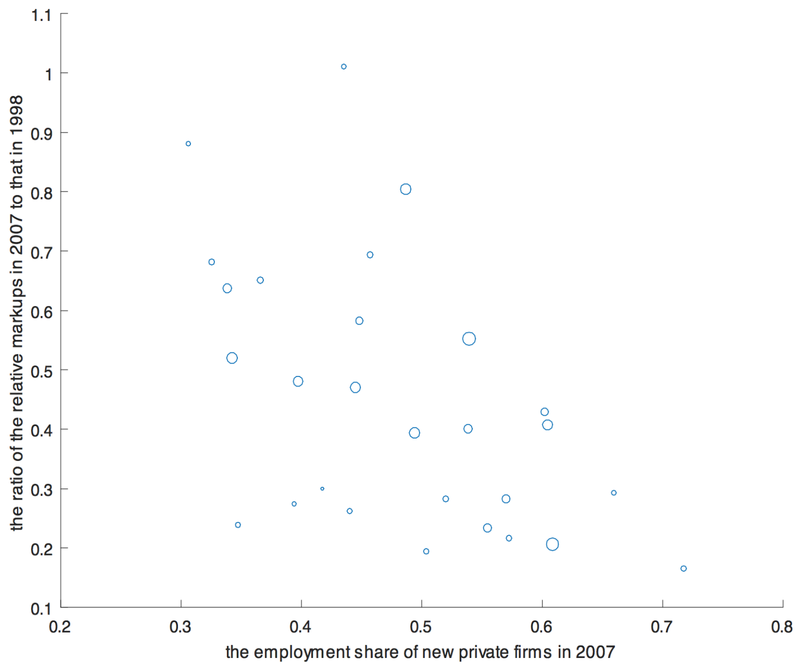

Figure 8: his figure plots the employment share of new private firms in manufacturing industries in 2007 (x-axis) and the ratio of the relative markups in that industry in 2007 to that in 1998 (y-axis).

deregulation would also increase the aggregate welfare by 30 percent, from which one fith is contributed from the variable markups channel.

There are certainly many other ways to think of variable markups within and across industries and to understand the underdevelopment of China's service sector. A major task to be done in the future is to show our mechanism and quantitative results are robust to alternative setups. Also, we want to explore the other channels through which China's service sector is underdeveloped.

References

[1] Arkolakis, C., A. Costinot, Dave Donaldson, and A. Rodriguez-Clare (2012): “The Elusive Pro-Competitive Effects of Trade,” Working Paper.

[2] Atkeson, Andrew, and Ariel Burstein (2008): “Pricing-to-Market, Trade Costs, and International Relative Prices,” American Economic Review, 98(5), 1998-2031.

[3] Auerbach, Alan J. (1985): ”The Theory of Excess Burden and Optimal Taxation,” in Handbook of Public Economics, Vol. 1. A.J. Auerbach and M. Feldstein, eds. Amsterdam: North Holland, 61-127.

[4] Brandt, Loren, Van Biesebroeck, Johannes and Zhang, Yifan, (2012): “Creative Accounting or Creative Destruction? Firm-level productivity growth in Chinese manufacturing,” Journal of Development Economics, 97(2), 339-351,

[5] Brandt, Loren, Johannes Von Biesebroeck and Yifan Zhang, (2015): “WTO and the Effect of Trade Liberalization on Productivity in Chinese Manufacturing”, Working Paper.

[6] Brandt, Loren, Gueorgui Kambourov and Kjetil Storesletten, (2016): “Firm Entry and Regional Growth Disparties: The Effects of SOEs in China,” Working Paper.

[7] Brandt, Loren, Trevor Tombe and Xiaodong Zhu, (2013): ”Factor Market Distortions Across Time, Space, and Sectors in China,” Review of Economic Dynamics, 16(1), 39-58.

[8] Brandt, Loren & Zhu, Xiaodong, 2010. ”Accounting for China’s Growth,” IZA Discussion Papers 4764.

[9] Bernard, A. B., J. Eaton, J. B. Jensen and S. Kortum, (2003): ”Plants and Productivity in International Trade,” American Economic Review, 93(4): 1268-1290.

[10] Behrens, Kristian and Yasusada Murata, (2012): “Trade, competition, and efficiency, Journal of International Economics,” 87(1), 1-17.

[11] Behrens Kristian, Giordano Mion, Yasusada Murata, and Jens Sudekum, (2014): “Trade, wages, and productivity,” International Economic Review, forthcoming.

[12] Bas, Maria, Thierry Mayer, and Mathias Thoenig, (2015): “From Micro to Macro: Demand, Supply, and Heterogeneity in the Trade Elasticity,” Working Paper

[13] Chaney Thomas, (2008): “Distorted Gravity: The Intensive and Extensive Margins of International Trade, ” American Economic Review, 98(4): 1707-1721

[14] De Blas Beatriz, and Katheryn Russ (2012): “Understanding Markups in the Open Economy under Bertrand Competition,” Working Paper.

[15] Dhingra, Swati and John Morrow, (2012): “The Impact of Integration on Productivity and Welfare Distortions Under Monopolistic Competition,” CEP Discussion Papers 1130.

[16] De Loecker, Jan and Frederic Warzynski, (2012): ”Markups and Firm-Level Export Status.” American Economic Review, 102(6), 2437-71.

[17] Edmond, Chris, Virgiliu Midrigan and Daniel Yi Xu (2015): ”Competition, Markups, and the Gains from International Trade,” American Economic Review, 105(10), 3183-3221.

[18] Eeckhout, Jan, (2004): ”Gibrat’s Law for (All) Cities.” American Economic Review, 94(5), 1429-1451.

[19] Epifani, Paolo and Gino Gancia, (2011): ”Trade, Markup Heterogeneity, and Misallocations,” Journal of International Economics, 83(1), 1-13.

[20] Feenstra Robert C. (2014): ”Restoring the Product Variety and Pro-competitive Gains from Trade with Heterogeneous Firms and Bounded Productivity,” NBER Working Paper 19833.

[21] Fernandes, A., P. Klenow, S. Meleshchuk, M. D. Pierola and U. A. Rodriguez-Clare, (2015): “The Intensive Margin Puzzle in Trade,” Working Paper.

[22] Head, Keith and Thierry Mayer, (2014): “Gravity Equations: Workhorse, Toolkit, and Cookbook, Chapter to appear in the Handbook of International Economics Vol. 4, eds. Gopinath, Helpman, and Rogoff.

[23] Herrendorf Berthold, Rogerson Richard, Valentinyi Akos (2014): ”Growth and Structural Transformation” in Handbook of Economic Growth, Vol. 2, Amsterdam: North Holland, 855–941.

[24] Holmes, Thomas J., Wen-Tai Hsu and Sanghoon Lee, (2014): “Allocative Efficiency, Mark-ups, and the Welfare Gains from Trade,” forthcoming in Journal of International Economics.

[25] Hsieh, Chang-Tai and Peter J. Klenow, (2009): ”Misallocation and Manufacturing TFP in China and India,” Quarterly Journal of Economics, 2009, 124 (4), 1403-1448.

[26] Hsieh, Chang-Tai and Zheng Song, (2015): “Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China,” Brookings Papers on Economic Activity, 50(1): 295-366.

[27] Hsu, Wen-Tai, Yi Lu and Guiying Wu, (2016): “Competition and Gains from Trade: A Quantitative Analysis of China Between 1995 and 2004,” Working Paper.

[28] Kichko, Sergey, Sergey Kokovin, Evgeny Zhelobodko, (2013): “Trade Patterns and Export Pricing Under non-CES Preferences”, Working Paper.

[29] Krugman, Paul R., (1979): ”Increasing Returns, Monopolistic Competition, and International Trade,” Journal of International Economics, 9(4), 469-479.

[30] Lerner, A. P., (1934): “The Concept of Monopoly and the Measurement of Monopoly Power,” Review of Economic Studies 1, 157–175.

[31] Lipsey, R. G. and Kelvin Lancaster, (1956): “The General Theory of Second Best,” Review of Economic Studies, 24(1): 11-32.

[32] Melitz, Marc J., (2003): “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” Econometrica, 71(6), 1695-1725.

[33] Melitz, Marc J. and G. Ottaviano, (2008): “Market Size, Trade, and Productivity,” Review of Economic Studies, 75(1): 295-316.

[34] Peters, Michael, (2012): “Heterogenous Mark-Ups and Endogenous Misallocation,” Working Paper.

[35] Mohring, Herbert, (1971): ”Alternative Welfare Gain and Loss Measures,” Western Economic Journal, 9, 349-68.

[36] Restuccia, D. and R. Rogerson, (2008): “Policy Distortions and Aggregate Productivity with Heterogeneous Plants,” Review of Economic Dynamics, 11, 707-20.

[37] Rutkowski, Ryan, (2015): “Service Sector Reform in China,” Peterson Institute for International Economics.

[38] Robinson, Joan, (1934): “The Economics of Imperfect Competition,” Macmillan.

[39] Slesnick, Daniel T., (1998): “Empirical Approaches to the Measurement of Welfare,” Journal of Economic Literature, 36(4), 2108–2165.

[40] Song, Zheng, Kjetil Storesletten and Fabrizio Zilibotti, (2011): "Growing Like China." American Economic Review, 101(1), 196-233.

[41] Song, Zheng, Duncan Thomas, Daniel Yi Xu and Miaojun Wang, (2016): "Manufacturing to Retail: Anatomy of Chinese Footwear Firms," Working Paper.

[42] Song, Zheng and Laura Guiying Wu, (2014): "Identifying Capital Misallocation," Working Paper.

[43] Zhelobodko, E., S. Kokovin, M. Parenti. and J. F. Thisse, (2012), "Monopolistic Competition: Beyond the Constant Elasticity of Substitution," Econometrica, 2012(80), 2765–2784.

Table 1: Firm Number and Size

	Number of Industries	Number of Firms	Mean of Employment (Person)
Manufacturing	29	362661	190.4
Service	34	549067	60.8

Note: The numbers are statistics of the sample truncated by the revenue scale of five million Yuan.

Table 2: Regressions on Firm Labor Productivity

VARIABLES	(1) log_yl	(2) log_yl	(3) log_yl	(4) log_yl
log_rev	0.300***	0.261***	0.278***	0.291***
	(0.00987)	(0.0101)	(0.0101)	(0.0104)
log_rev_SD	0.154***	0.180***	0.175***	0.172***
	(0.0475)	(0.0498)	(0.0487)	(0.0481)
log_kl		0.255***	0.259***	0.261***
		(0.00806)	(0.00801)	(0.00806)
log_kl_SD		-0.104***	-0.0981***	-0.100***
		(0.0128)	(0.0122)	(0.0124)
soe			-0.408***	0.552***
			(0.0410)	(0.122)
age			-0.0113***	-0.0112***
			(0.00173)	(0.00170)
log_rev_soe				-0.114***
				(0.0175)
log_kl_soe				0.00362
				(0.0112)
Industry Dummies	+	+	+	+
Observations	2,721,726	2,498,610	2,497,651	2,497,651
R-squared	0.572	0.608	0.620	0.622

Note: log_yl, log_rev and log_kl stand for log labor productivity, log revenue and the log of the capital-labor ratio, respectively. SD is a dummy variable that equals one for service industries. soe is also a dummy variable that equals one for SOEs. age is the age of a firm since it is established. log_rev_SD and log_kl_SD are the interaction terms between log_rev, log_kl and SD. log_rev_soe and log_kl_soe are the interaction terms between log_rev, log_kl and soe. Industry Dummies are for two-digit industries. Standard errors are clustered at the industry level and reported in parentheses. ***, ** and * stand for the statistical significance level at 1%, 5% and 10%, respectively.

Table 3: Calibration Results

	rho	cd	s.d.	R^2 of the markups profile	R^2 square of the sales profile
Processing of Food from Agricultural Products	0.26	0.65	4.06	0.96	0.89
Manufacture of Foods	0.40	0.53	2.85	0.99	0.99
Manufacture of Beverages	0.44	0.60	2.60	0.99	1.00
Manufacture of Textile	0.41	0.72	2.30	0.99	0.99
Manufacture of Textile Wearing Apparel, Footware and Caps	0.42	0.50	2.09	0.98	0.97
Manufacture of Leather, Fur, Feather and Related Products	0.51	0.75	1.59	0.99	1.00
Processing of Timber, Manufacture of Wood, Bamboo, Rattan,Palm and Straw Products	0.37	0.50	2.36	0.99	0.98
Manufacture of Furniture	0.51	0.76	1.49	1.00	1.00
Manufacture of Paper and Paper Products	0.34	0.58	3.40	0.93	0.97
Printing, Reproduction of Recording Media	0.55	0.54	1.30	0.99	1.00
Manufacture of Articles For Culture, Education and Sport Activities	0.60	0.67	1.02	1.00	1.00
Processing of Petroleum, Coking, Processing of Nuclear Fuel	0.17	0.50	0.21	0.48	0.39
Manufacture of Raw Chemical Materials and Chemical Products	0.36	0.70	3.68	0.99	1.00
Manufacture of Medicines	0.41	0.53	2.64	1.00	1.00
Manufacture of Chemical Fibers	0.29	0.55	5.70	0.92	1.00
Manufacture of Rubber	0.36	0.62	3.54	0.86	0.97
Manufacture of Plastics	0.49	0.59	1.66	0.99	1.00
Manufacture of Non-metallic Mineral Products	0.33	0.50	3.15	0.98	0.99
Smelting and Pressing of Ferrous Metals	0.32	0.53	5.35	0.95	0.99
Smelting and Pressing of Non-ferrous Metals	0.30	0.56	5.35	0.95	1.00
Manufacture of Metal Products	0.41	0.50	2.45	0.98	0.99
Manufacture of General Purpose Machinery	0.45	0.51	2.27	0.99	0.99
Manufacture of Special Purpose Machinery	0.45	0.50	2.37	0.98	0.99
Manufacture of Transport Equipment	0.30	0.67	5.28	0.75	0.98
Manufacture of Electrical Machinery and Equipment	0.38	0.66	3.39	0.93	1.00
Manufacture of Communication Equipment, Computers and Other Electronic Equipment	0.32	0.90	5.49	0.70	1.00
Manufacture of Measuring Instruments and Machinery for Cultural Activity and Office Work	0.49	0.51	2.17	0.88	0.99
Manufacture of Artwork and Other Manufacturing	0.37	0.50	2.55	0.89	0.94
Recycling and Disposal of Waste	0.20	0.50	2.49	0.96	0.68
Road Transport	0.64	0.51	0.96	0.99	0.98

Water Transport	0.49	0.53	2.58	0.94	0.99
Loading, Unloading and Other Transport Services	0.44	0.50	2.39	0.92	0.97
Storage	0.19	0.59	1.63	0.92	0.45
Telecommunications and Other Information Transmission Services	0.36	0.50	4.61	0.73	0.96
Computer Services	0.66	0.52	1.15	0.96	0.99
Software	0.78	0.65	0.51	0.99	1.00
Wholesale Trade	0.29	0.69	3.92	0.82	0.79
Retail Trade	0.46	0.50	2.41	0.97	1.00
Hotels	0.81	2.35	0.20	1.00	1.00
Catering Services	0.79	1.12	0.21	1.00	0.90
Real Estate	0.15	0.53	3.79	0.97	0.60
Leasing	0.59	0.54	1.00	0.96	0.96
Business Services	0.60	0.73	1.42	0.92	0.98
Research and Experimental Development	0.83	0.59	1.22	0.96	0.99
Professional Technical Services	0.67	0.64	1.03	0.95	0.98
Services of Science and Technology Exchanges and Promotion	0.67	0.64	1.01	0.92	0.98
Geologic Prospecting	0.69	0.52	1.30	0.97	1.00
Management of Water Conservancy	0.45	0.58	1.51	0.94	0.94
Environmental Management	0.58	0.58	0.98	0.94	0.97
Management of Public Facilities	0.48	0.61	1.52	0.98	0.98
Services to Households	0.72	0.65	0.42	0.99	0.98
Education	0.82	0.50	0.27	0.99	0.99
Health	0.78	0.50	0.46	0.99	0.99
Journalism and Publishing Activities	0.59	0.51	1.25	0.99	0.99
Broadcasting, Movies, Television and Audiovisual Activities	0.65	0.60	1.10	0.99	0.98
Entertainment	0.74	0.54	0.39	0.98	0.97

8 Online Appendix

The appendix contains outlines of the proofs of the propositions.

8.1 Proof of Lemma 1

First, Let us prove that under Assumption 1, q_i, markups, revenue and profits are all strictly decreasing in c.

Under the Assumption 1, it is trivial to show that markups are decreasing in c.

With the help of assumption 1, we obtain that

$$\frac{\partial \left[u'(q)(1-\mu(q))\right]}{\partial q} = u''(q)(1-\mu(q)) - u'(q)\mu'(q) < 0,$$

i.e., $u'(q)(1-\mu(q))$ is a decreasing function of q. From the following equation,

$$u'(q(c))(1-\mu(q(c))) = u'(0)(1-\mu(0))\frac{c}{c_D},$$

we obtain that

$$\frac{\partial \left[u'(q(c))(1-\mu(q(c)))\right]}{\partial c} = \frac{\partial \left[u'(q(c))(1-\mu(q(c)))\right]}{\partial q}\frac{\partial q}{\partial c} = \frac{u'(0)(1-\mu(0))}{c_D} \Rightarrow$$

$$\frac{\partial q}{\partial c} < 0,$$

i.e., q is a decreasing function of c. Also, we can obtain that

$$p(c)\,q(c) = \frac{c}{(1-\mu(q(c)))}q(c) = \frac{c_D}{u'(0)(1-\mu(0))}u'(q(c)).$$

Since $u(q)$ is a concave function, we know that revenue $p(c)\,q(c)$ is a decreasing function of c. It is easy to see that profit

$$\pi(c) = \left[\frac{1}{1-\mu(q(c))} - 1\right]p(c)\,q(c)$$

is also strictly decreasing in c.

8.2 Proof of Proposition 1

The household problem is

$$\max \int_0^N u(q_i)\,di,$$

subject to $\int_0^N p_i q_i di = I$. The first order condition,

$$p_i = \frac{u'(q_i)}{\lambda},$$

gives us the demand function for each variaty. Given the demand function, the firm's problem is

$$\max pq - cq.$$

The first order condition is

$$u'(q)(1-\mu(q)) = c.$$

At the cutoff, $q = 0$. Hence, we have the following key equation:

$$u'(q)(1-\mu(q)) = u'(0)(1-\mu(0))\frac{c}{c_D}.$$

From this equation, we know that q is a function of $\frac{c}{c_D}$.

$\bar{N}$ potential firms are going to enter the market and their marginal cost follows the distribution characterized by $G(c)$. Then, the measure and productivity distribution of the active firms are $N = \bar{N}G(c_D)$ and $F(c) = \frac{G(c)}{G(c_D)}$, respectively. We can now write down the expression for the aggregate utility, revenue and employment:

$$U = N\int_0^{c_D} u\left[q\left(\frac{c}{c_D}\right)\right] dF(c),$$

$$R = N\int_0^{c_D} p(c)\, q\left(\frac{c}{c_D}\right) dF(c) = N\int_0^{c_D} \frac{c}{1-\mu\left(q\left(\frac{c}{c_D}\right)\right)} q\left(\frac{c}{c_D}\right) dF(c),$$

$$L = N\int_0^{c_D} cq\left(\frac{c}{c_D}\right) dF(c).$$

Define $x = c/c_D$, from the key equation, we know that q can be expressed as a function x. Then, we get the following equations:

$$\begin{aligned} L &= Nc_D\int_0^{c_D} \frac{c}{c_D} q\left(\frac{c}{c_D}\right) dF(c) \\ &= \bar{N}G(c_D)\, c_D M\int_0^1 xq(x)\, d\Psi(x) \\ &= \bar{N}G(c_D)\, c_D X, \end{aligned}$$

where $X = \int_0^1 xq(x)\, d\Psi(x)$;

$$\begin{aligned} R &= N\int_0^{c_D} p(c)\, q(c)\, dF(c) \\ &= c_D\bar{N}G(c_D)\int_0^1 \frac{xq(x)}{1-\mu(q(x))} d\Psi(x) \\ &= \bar{N}G(c_D)\, c_D Y, \end{aligned}$$

where $Y = \int_0^1 \frac{xq(x)}{1-\mu(q(x))} d\Psi(x)$;

$$\begin{aligned} U &= M \int_0^1 u[q(x)]\, d\Psi(x) \\ &= MZ, \end{aligned}$$

where $Z = \int_0^1 u[q(x)]\, d\Psi(x)$. Finally, the average markups are

$$\Phi = \frac{R}{L} = \frac{\bar{N}G(c_D)\, c_D Y}{\bar{N}G(c_D)\, c_D X} = \frac{Y}{X}.$$

To prove the averge markups associated with Ψ^2 are higher than that associated with Ψ^1, what we need to do is to prove that $\frac{Y^2}{X^2}$ associated with Ψ^2 is higher than that associated with Ψ^1. Let $y = \phi(x) \equiv (\Psi^1)^{-1}(\Psi^2(x))$. Since Ψ^1 first-order stochastically dominates Ψ^2, i.e., $\Psi^1(x) \le \Psi^2(x)$ for all x, we have

$$y = \phi(x) \ge x.$$

We can rewrite Y^1 and X^1 as

$$\begin{aligned} Y^1 &= \int_0^1 \frac{xq(x)}{1-\mu(q(x))} d\Psi^1(x) \\ &= \int_0^1 \frac{yq(y)}{1-\mu(q(y))} d\Psi^1(y) \\ &= \int_0^1 \frac{\phi(x)\, q(\phi(x))}{1-\mu(q(\phi(x)))} d\Psi^2(x), \end{aligned}$$

and

$$\begin{aligned} X^1 &= \int_0^1 xq(x)\, d\Psi^1(x) = \int_0^1 yq(y)\, d\Psi^1(y) \\ &= \int_0^1 \phi(x)\, q(\phi(x))\, d\Psi^2(x). \end{aligned}$$

For every x, we know that

$$q(\phi(x)) \le q(x),$$

and

$$\frac{1}{1-\mu(q(\phi(x)))} \le \frac{1}{1-\mu(q(x))}.$$

This leads to the conclusion that

$$Y^1 \le Y^2 \text{ and } Z^1 \le Z^2.$$

For every x, we have

$$\left[\frac{xq(x)}{1-\mu(q(x))}\right] / \left[\frac{\phi(x)\, q(\phi(x))}{1-\mu(q(\phi(x)))}\right] \ge [xq(x)] / [\phi(x)\, q(\phi(x))].$$

It implies that

$$\frac{\int_0^1 \frac{xq(x)}{1-\mu(q(x))} d\Psi^2(x)}{\int_0^1 \frac{\phi(x)q(\phi(x))}{1-\mu(q(\phi(x)))} d\Psi^2(x)} \geq \frac{\int_0^1 xq(x)\, d\Psi^2(x)}{\int_0^1 \phi(x)\, q(\phi(x))\, d\Psi^2(x)},$$

$$\frac{Y^2}{Y^1} \geq \frac{X^2}{X^1} \Rightarrow \frac{Y^2}{X^2} \geq \frac{Y^1}{X^1}.$$

We also need to prove that $Z^2/X^2 \geq Z^1/X^1$. This claim is going to be used in the proof of proposition 3.

Let us prove that $\partial \frac{u(q)}{xq} / \partial x < 0$. From the first order conditon,

$$u'(q)(1-\mu(q)) = u'(0)(1-\mu(0)) \frac{c}{c_D},$$

we know that

$$xq = \frac{u'(q)(1-\mu(q))\, q}{u'(0)(1-\mu(0))}.$$

To prove that $\partial \frac{u(q)}{xq} / \partial x < 0$, it is equavelant to prove that

$$\frac{\partial \left[\frac{u(q)}{u'(q)(1-\mu(q))q} \right]}{\partial q} \frac{\partial q}{\partial x} < 0.$$

Since $\frac{\partial q}{\partial x} < 0$, we need to show that

$$\frac{\partial \left[\frac{u(q)}{u'(q)(1-\mu(q))q} \right]}{\partial q} > 0.$$

By assumption, we already know that $\frac{\partial \frac{1}{1-\mu(q)}}{\partial q} > 0$. So, it is sufficient to prove that $\frac{\partial \left[\frac{u(q)}{u'(q)q} \right]}{\partial q} > 0$. It is easy to prove that

$$\frac{\partial u(q)}{\partial q} > \frac{\partial u'(q)\, q}{\partial q}.$$

So, $\frac{\partial \left[\frac{u(q)}{u'(q)q} \right]}{\partial q} > 0$ holds true.

Now we have

$$\frac{u(q(x))}{xq(x)} \geq \frac{u(q(\phi(x)))}{\phi(x)\, q(\phi(x))}.$$

It implies that

$$\frac{\int_0^1 u(q(x))\, d\Psi^2(x)}{\int_0^1 xq(x)\, d\Psi^2(x)} \geq \frac{\int_0^1 u(q(\phi(x)))\, d\Psi^2(x)}{\int_0^1 \phi(x)\, q(\phi(x))\, d\Psi^2(x)},$$

which is

$$\frac{Z^2}{X^2} \geq \frac{Z^1}{X^1}.$$

Next, let us prove that Y/L distribution associated with $\Psi^1(x)$ is Lorenze dominated by Y/L distribution associated with $\Psi^2(x)$. The Lorenz curves of Y/L, Θ^1 and Θ^2, are

$$\Theta^i = \frac{\int_p^1 \frac{1}{1-\mu(q(\Psi^{-1,i}(p)))}dp}{\int_0^1 \frac{1}{1-\mu(q(\Psi^{-1,i}(p)))}dp},$$

where $i \in \{1,2\}$ and p is the probability quantile. $\Psi^{-1,i}(p)$ represents the inverse function of $\Psi^i(p)$. Since

$$\Psi^{-1,2}(p) \leq \Psi^{-1,1}(p),$$

$q\left(\Psi^{-1,2}(p)\right) \geq q\left(\Psi^{-1,1}(p)\right)$ for all value of $p \in [0,1]$. The ratio of the Lorenze curve associated with Ψ^1 and Ψ^2 is

$$\frac{\Theta^1(p)}{\Theta^2(p)} = \frac{\int_0^1 \frac{1}{1-\mu(q(\Psi^{-1,2}(p)))}dp \int_p^1 \frac{1}{1-\mu(q(\Psi^{-1,1}(p)))}dp}{\int_0^1 \frac{1}{1-\mu(q(\Psi^{-1,1}(p)))}dp \int_p^1 \frac{1}{1-\mu(q(\Psi^{-1,2}(p)))}dp}.$$

Since $q\left(\Psi^{-1,2}(p)\right) \geq q\left(\Psi^{-1,1}(p)\right)$, we can reach the conclusion that

$$\frac{\int_0^1 \frac{1}{1-\mu(q(\Psi^{-1,2}(p)))}dp}{\int_0^1 \frac{1}{1-\mu(q(\Psi^{-1,1}(p)))}dp} \geq \frac{\int_p^1 \frac{1}{1-\mu(q(\Psi^{-1,2}(p)))}dp}{\int_p^1 \frac{1}{1-\mu(q(\Psi^{-1,1}(p)))}dp} \geq 1 \Rightarrow$$

$$\frac{\Theta^1(p)}{\Theta^2(p)} \geq 1.$$

This means that the Lorenze curve associated with Ψ^1 is above that associated with Ψ^2. Therefore, the revenue-labor ratio distribution associated with Ψ^2 is more dispersed than that associated with Ψ^1.

8.3 Proof of Proposition 2

We take two steps to prove the first part of the proposition 2. The first step is to prove that if $\bar{N}$ increase, c_D decrease. Next, we can invoke Proposition 1 to establish the claim directly.

We first prove that c_D decreases in $\bar{N}$. The labor market clear condition gives

$$L = \bar{N}G(c_D)\int_0^{c_D} cq\left(\frac{c}{c_D}\right)d\frac{G(c)}{G(c_D)} = 1 \Rightarrow$$

$$\bar{N}\int_0^{c_D} cq\left(\frac{c}{c_D}\right)dG(c) = 1.$$

We know that if $\bar{N}$ increases, c_D will have to decrease to keep the labor market clear. This is the standard selection effect and does not hinge on the specific property of the productivity distribution.

We now discuss the case associated with Pareto distribution. The relative productivity distribution, $\Psi(x) = x^{\kappa}$, becomes independent of c_D. Therefore, a higher $\bar{N}$ reduces c_D but leaves $\Psi(c/c_D)$ unchanged. In turn, the invariant $\Psi(c/c_D)$ keeps the average markups and revenue-labor ratio dispersion unchanged.

Turn to the second part of proposition 2. Assume that firms face idiosyncratic distortions. To smplify the analysis, we adopt a binary wedge setup. More specificly, some firms face a positive tax rate, τ, and some firms face a negative tax rate, $-\tau$.

First consider the Pareto productivity distribution, $G(c) = \left(\frac{c}{\bar{c}}\right)^{\kappa}$. The number of potential firms is $\bar{N}$. Given the demand function, $p_i = u'(q_i)/\lambda$, firms maximize their profit:

$$\max_{q_i} (1-\tau) p_i q_i - c_i q_i.$$

The first-order condition is

$$\frac{[u''(q_i) q_i + u'(q_i)]}{\lambda} = \frac{c_i}{1-\tau},$$

which can be rewritten as

$$u'(q(c))(1-\mu(q(c))) = \frac{c/(1-\tau)}{c_{\tau,D}/(1-\tau)} = \frac{c}{c_{\tau,D}}.$$

The cutoff productivity is also binary and must satisfy

$$\frac{c_D^{+}}{1-\tau} = \frac{c_D^{-}}{1+\tau}.$$

We use the superscripts of $+$ and $-$ to represent the firms associated with τ and $-\tau$, respectively. The sales of the two types of firms follow

$$\begin{aligned}
R^{+} = p^{+}q^{+} = S\bar{N}\left(\frac{c_D^{+}}{\bar{c}}\right)^{\kappa} & \int_0^{c_D^{+}} \frac{c}{1-\tau} \frac{q\left(\frac{c}{c_D^{+}}\right)}{\left(1-\mu\left(q\left(\frac{c}{c_D^{+}}\right)\right)\right)} d\left(\frac{c}{c_D^{+}}\right)^{\kappa} \\
= S\bar{N}\left(\frac{c_D^{+}}{\bar{c}}\right)^{\kappa} & \frac{c_D^{+}}{1-\tau} \int_0^{c_D^{+}} \frac{c}{c_D^{+}} \frac{q\left(\frac{c}{c_D^{+}}\right)}{\left(1-\mu\left(q\left(\frac{c}{c_D^{+}}\right)\right)\right)} d\left(\frac{c}{c_D^{+}}\right)^{\kappa} \\
= S\bar{N}\left(\frac{c_D^{+}}{\bar{c}}\right)^{\kappa} & \frac{c_D^{+}}{1-\tau} \int_0^{1} x \frac{q(x)}{(1-\mu(q(x)))} d(x)^{\kappa} \\
= S\bar{N}\left(\frac{c_D^{+}}{\bar{c}}\right)^{\kappa} & \frac{c_D^{+}}{1-\tau} Y,
\end{aligned}$$

$$R^{-} = (1-S)\bar{N}\left(\frac{c_D^{-}}{\bar{c}}\right)^{\kappa} \frac{c_D^{-}}{1+\tau} Y.$$

Their after-tax sales revenue are

$$
\begin{aligned}
AR^{+} &= (1-\tau)\, R^{+} \\
&= S\bar{N}\left(\frac{c_D^{+}}{\bar{c}}\right)^{\kappa} c_D^{+} Y,
\end{aligned}
$$

$$
\begin{aligned}
AR^{-} &= (1+\tau)\, R^{-} \\
&= (1-S)\left(\frac{c_D^{-}}{\bar{c}}\right)^{\kappa} c_D^{-} Y.
\end{aligned}
$$

Employment follows

$$
\begin{aligned}
L^{+} &= S\bar{N}\left(\frac{c_D^{+}}{\bar{c}}\right)^{\kappa} \int_0^{c_D^{+}} cq\left(\frac{c}{c_D^{+}}\right) d\left(\frac{c}{c_D^{+}}\right)^{\kappa} \\
&= S\bar{N}\left(\frac{c_D^{+}}{\bar{c}}\right)^{\kappa} c_D^{+} \int_0^{1} xq\,(x)\, d\,(x)^{\kappa} \\
&= S\bar{N}\left(\frac{c_D^{+}}{\bar{c}}\right)^{\kappa} c_D^{+} X,
\end{aligned}
$$

$$
L^{-} = (1-S)\, \bar{N}\left(\frac{c_D^{-}}{\bar{c}}\right)^{\kappa} c_D^{-} X.
$$

We introduce the balanced-budget constraint for τ:

$$
R^{+} + R^{-} = AR^{+} + AR^{-}.
$$

The idea is that if we literally interpret τ and $-\tau$ as taxes and subsidies, the government would have to run a balance budget for each industry.

Then average markups are

$$
\frac{R}{L} = \frac{AR}{L} = \frac{\left(c_D^{+}\right)^{\kappa+1} + \left(c_D^{-}\right)^{\kappa+1}}{\left(c_D^{+}\right)^{\kappa+1} + \left(c_D^{-}\right)^{\kappa+1}} \frac{Y}{X} = \frac{Y}{X}.
$$

Obviously, Y/L is constant regardless of τ.

We now move away from the Pareto distribution and prove the claim under Assumption 2. Let us first characterize the equilibrium without idiosyncratic distortions. Let c_D^{*} be the cutoff. The productivity distribution is defined as $G^{*}\,(c)$ and the corresponding transformed distribution function is $\Psi^*(x)$. Then, the sales of firms are

$$\begin{aligned}
R^* = pq &= \bar{N}G(c_D^*)\int_0^{c_D^*} c\frac{q\left(\frac{c}{c_D^*}\right)}{\left(1-\mu\left(q\left(\frac{c}{c_D^*}\right)\right)\right)} d\frac{G^*(c)}{G(c_D^*)} \\
&= \bar{N}G(c_D^*)\, c_D^* \int_0^{c_D^*} \frac{c}{c_D^*}\frac{q\left(\frac{c}{c_D^*}\right)}{\left(1-\mu\left(q\left(\frac{c}{c_D^*}\right)\right)\right)} dG^*(c) \\
&= \bar{N}G(c_D^*)\, c_D^* \int_0^1 x\frac{q(x)}{(1-\mu(q(x)))} d\Psi^*(x) \\
&= \bar{N}G(c_D^*)\, c_D^* Y^*.
\end{aligned}$$

Employment is

$$\begin{aligned}
L^* &= \bar{N}G(c_D^*)\int_0^{c_D^*} cq\left(\frac{c}{c_D^*}\right) dG^*(c) \\
&= \bar{N}G(c_D^*)\, c_D^* \int_0^1 xq(x)\, d\Psi^*(x) \\
&= \bar{N}G(c_D^*)\, c_D^* X^*.
\end{aligned}$$

Then

$$\frac{R^*}{L^*} = \frac{Y^*}{X^*}.$$

When idiosyncratic distortions are present, we have

$$\begin{aligned}
R^+ = p^+q^+ &= S\bar{N}G\left(c_D^+\right)\int_0^{c_D^+} \frac{c}{1-\tau}\frac{q\left(\frac{c}{c_D^+}\right)}{\left(1-\mu\left(q\left(\frac{c}{c_D^+}\right)\right)\right)} dF^+(c) \\
&= S\bar{N}G\left(c_D^+\right)\frac{c_D^+}{1-\tau}\int_0^{c_D^+} \frac{c}{c_D^+}\frac{q\left(\frac{c}{c_D^+}\right)}{\left(1-\mu\left(q\left(\frac{c}{c_D^+}\right)\right)\right)} dF^+(c) \\
&= S\bar{N}G\left(c_D^+\right)\frac{c_D^+}{1-\tau}\int_0^1 x\frac{q(x)}{(1-\mu(q(x)))} d\Psi^+(c) \\
&= S\bar{N}G\left(c_D^+\right)\frac{c_D^+}{1-\tau} Y^+,
\end{aligned}$$

$$\begin{aligned}
R^- &= (1-S)\,\bar{N}G\left(c_D^-\right)\frac{c_D^-}{1+\tau}\int_0^1 x\frac{q(x)}{(1-\mu(q(x)))} d\Psi^-(c) \\
&= (1-S)\,\bar{N}G\left(c_D^-\right)\frac{c_D^-}{1+\tau} Y^-,
\end{aligned}$$

$$\begin{aligned} AR^{+} &= (1-\tau)\,R^{+} \\ &= S\bar{N}G\left(c_D^{+}\right)c_D^{+}Y^{+}, \end{aligned}$$

$$\begin{aligned} AR^{-} &= (1+\tau)\,R^{-} \\ &= (1-S)\,\bar{N}G\left(c_D^{-}\right)c_D^{-}Y^{-}, \end{aligned}$$

$$\begin{aligned} L^{+} &= S\bar{N}G\left(c_D^{+}\right)c_D^{+}\int_0^1 xq\,(x)\,d\Psi^{+}\,(c) \\ &= S\bar{N}G\left(c_D^{+}\right)c_D^{+}X^{+}, \end{aligned}$$

$$L^{-} = (1-S)\,\bar{N}G\left(c_D^{-}\right)c_D^{-}X^{-}.$$

Once again, the budget must be balanced, i.e., $R^{+} = R^{-}$ or $\frac{AR^{+}}{1-\tau} = \frac{AR^{-}}{1+\tau}$. Then, the average markups are

$$\frac{Y}{L} = \frac{AR^{+}+AR^{-}}{L^{+}+L^{-}} = \frac{SG\left(c_D^{+}\right)c_D^{+}Y^{+} + (1-S)\,G\left(c_D^{-}\right)c_D^{-}Y^{-}}{SG\left(c_D^{+}\right)c_D^{+}X^{+} + (1-S)\,G\left(c_D^{-}\right)c_D^{-}X^{-}}.$$

Since the aggregate labor is normalized into unity, we obtain that

$$\begin{gathered} SG\left(c_D^{+}\right)c_D^{+}X^{+} + (1-S)\,G\left(c_D^{-}\right)c_D^{-}X^{-} = \\ SG\left(c_D^{*}\right)c_D^{*}X^{*} = \bar{L}. \end{gathered}$$

Since $c_D^{-} > c_D^{*} > c_D^{+}$, from the previous proof of the first part of Proposition 2, we know that

$$X^{-} > X^{*} > X^{+}, Y^{-} > Y^{*} > Y^{+},$$

and

$$\frac{Y^{-}}{X^{-}} > \frac{Y^{*}}{X^{*}} > \frac{Y^{+}}{X^{+}}.$$

The labor market clear condition implies

$$S\bar{N}G\left(c_D^{+}\right)c_{+,D}X^{+} + (1-S)\,\bar{N}G\left(c_D^{-}\right)c_D^{-}X^{-} = \bar{N}G\left(c_D^{*}\right)c_D^{*}X^{*} = \bar{L}.$$

This gives

$$\frac{\partial\left[S\bar{N}G\left(c_D^{+}\right)c_D^{+}X^{+}\right]}{\partial\tau} + \frac{\partial\left[(1-S)\,\bar{N}G\left(c_D^{-}\right)c_D^{-}X^{-}\right]}{\partial\tau} = 0.$$

To prove that the average markup increases, it will be equivalent to prove that

$$\frac{\partial R}{\partial \tau} > 0.$$

We know that

$$\begin{aligned}\frac{\partial R}{\partial \tau} &= \frac{\partial\left[S\bar{N}G\left(c_D^+\right)c_D^+\Phi^+X^+ + (1-S)\,\bar{N}G\left(c_D^-\right)c_D^-\Phi^-X^-\right]}{\partial \tau} \\ &= \left\{\frac{\partial\left[S\bar{N}G\left(c_D^+\right)c_D^+X^+\right]}{\partial \tau}\Phi^+ + \frac{\partial\left[(1-S)\,\bar{N}G\left(c_D^-\right)c_D^-X^-\right]}{\partial \tau}\Phi^-\right\} \\ &+ \left\{S\bar{N}G\left(c_D^+\right)c_D^+X^+\frac{\partial \Phi^+}{\partial \tau} + \left[(1-S)\,\bar{N}G\left(c_D^-\right)c_D^-X^-\right]\frac{\partial \Phi^-}{\partial \tau}\right\}\end{aligned}$$

Since $\Phi^- > \Phi^+$, it is easy to see that the first term on the right hand side of equation is positive. When τ is small, we know that

$$\frac{\partial \Phi^-}{\partial \tau} = -\frac{\partial \Phi^+}{\partial \tau} > 0.$$

Moreover, we have already known that $(1-S)\,\bar{N}G\left(c_D^-\right)c_D^-X^- > S\bar{N}G\left(c_D^+\right)c_D^+X^+$. So, we conculde that the second term on the right hand side of equation is positive. In sum, $\frac{\partial R}{\partial \tau} > 0$ and we know that average markups, Y/L, increase with the magnitude of the idiosyncratic distortions.

8.4 Proof of Proposition 3

The household problem is

$$\max\left[\sum_{j}^{2}\gamma^i U^{j\,\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}}$$

$$s.t\ \int_0^{N^1} p_i^1 q_i^1 di + \int_0^{N^2} p_i^2 q_i^2 di = I.$$

Solving the household problem gives the inverse demand function:

$$p_i^j = \gamma^j\left[\sum_{j}^{2}\gamma^j U^{j\,\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}} U^{j-1/\sigma}u'\left(q_i^j\right)/\lambda,$$

Within a sector, the first-order condition implies

$$\frac{u'\left(q^j(c)\right)\left(1-\mu\left(q^j(c)\right)\right)}{u'\left(q^j\left(c_D^j\right)\right)\left(1-\mu\left(q^j\left(c_D^j\right)\right)\right)} = \frac{c}{c_D^j}.$$

Upon entry, each firm receives a unit cost c drawn from a Pareto distribution $G(c)$.

The cutoff productivity is c_D^j. We know that $q\left(c_D^j\right) = 0$ and $p\left(c_D^j\right) = c_D^j$. Therefore, the first-order condition can be rewritten as

$$u'\left(q^j(c)\right)\left(1-\mu\left(q^j(c)\right)\right) = u'(0)\,\frac{c}{c_D^j}.$$

By the demand function, we can also obtain

$$p\left(c_D^j\right) = \gamma^j \left[\sum_j^2 U^{j\frac{\sigma-1}{\sigma}}\right]^{\frac{1}{\sigma-1}} U^{j-1/\sigma} u'\left(q\left(c_D^j\right)\right)/\lambda = c_D^j.$$

This leads to the condition governing the cutoff productivity across sectors:

$$\frac{c_D^j}{c_D^{j'}} = \frac{\gamma^j}{\gamma^{j'}}\left(\frac{U^j}{U^{j'}}\right)^{-1/\sigma}. \tag{12}$$

In equilibrium, the number of varieties, utility, average expenditure and employment in each sector will follow

$$N^j = \bar{N}^j G\left(c_D^j\right),$$

$$\begin{aligned} U^j &= N_i \int_0^{c_D^j} u\left[q\left(\frac{c}{c_D^j}\right)\right] dF^j(c) \\ &= N^j \int_0^1 u\left[q(x)\right] d\Psi^j(x) \\ &= \bar{N}^j G\left(c_D^j\right) Z^j, \end{aligned}$$

$$\begin{aligned} R^j &= N^j \int_0^{c_D^j} p(c)\, q\left(\frac{c}{c_D^j}\right) dF^j(c) \\ &= \bar{N}^j G\left(c_D^j\right) c_D^j \int_0^1 \frac{x}{1-\mu\left(q(x)\right)} q(x)\, d\Psi^j(x) \\ &= \bar{N}^j G\left(c_D^j\right) c_D^j Y^j, \end{aligned}$$

$$\begin{aligned} L^j &= N^j \int_0^{c_D^j} cq\left(\frac{c}{c_D^j}\right) dF^j(c) \\ &= \bar{N}^j G\left(c_D^j\right) c_D^j \int_0^1 xq(x)\, d\Psi^j(x) \\ &= \bar{N}^j G\left(c_D^j\right) c_D^j X^j. \end{aligned}$$

Rewrite (12) as

$$\frac{c_D^j}{c_D^{j'}} = \frac{\gamma^j}{\gamma^{j'}}\left(\frac{U^j}{U^{j'}}\right)^{-1/\sigma} \Rightarrow \frac{c_D^j}{c_D^{j'}} = \frac{\gamma^j}{\gamma^{j'}}\left(\frac{\bar{N}^j G\left(c_D^j\right) Z^j}{\bar{N}^{j'} G\left(c_D^{j'}\right) Z^{j'}}\right)^{-1/\sigma} \Rightarrow$$

$$\frac{\bar{N}^j G\left(c_D^j\right) Z^j \left(c_D^j\right)^\sigma}{\bar{N}^{j'} G\left(c_D^{j'}\right) Z^j \left(c_D^{j'}\right)^\sigma} = \frac{\gamma^j}{\gamma^{j'}} \Rightarrow \frac{L^j}{L^{j'}} = \frac{\gamma^j}{\gamma^{j'}} \left(\frac{c_D^j}{c_D^{j'}}\right)^{1-\sigma} \left(\frac{Z^j/X^j}{Z^{j'}/X^{j'}}\right)^{-1}. \tag{13}$$

The second key equation is derived from the labor market clear condition

$$L^1 + L^2 = \bar{N}^1 G\left(c_D^1\right) c_D^1 X^1 + \bar{N}^2 G\left(c_D^2\right) c_D^2 X^2 = L = 1.$$

Now we are well eqipped to discuss the comparative statics of reducing entry barrier in sector 1, i.e., increasing $\bar{N}^1$. First, we consider the case in which Assumption 2 holds. We already proved X^j, Z^j and Z^j/X^j all are an increasing function of c_D^j. Let us consider the effect of $\bar{N}^1$ on c_D^2 under the condition $\sigma \geq 1$. If c_D^2 does not change, the labor market clear condition will imply that L^j keeps unchanged and c_D^1 has to decrease. But it contradicts the condition (13). If c_D^2 increases, the labor market clear condition will imply that L^2 increases and c_D^1 has to decrease. Again, it contradicts the condition (13). Hence, c_D^2 and L^2 must decrease.

Turn to the case of $\sigma > 1$. Following exactly the same logic above, one can easily show that c_D^2 and L^2 must decrease.

If the productivity distribution is Pareto, $G^j(c) = \left(\frac{c}{\bar{c}^j}\right)^{\kappa^j}$, we will get

$$\gamma^2 \left(c_D^1\right)^{1+\kappa^1/\sigma} \left(\bar{N}^1 \left(\bar{c}^1\right)^{-\kappa^1} Z^1\right)^{1/\sigma} = \gamma^1 \left(c_D^2\right)^{1+\kappa^2/\sigma} \left(\bar{N}^2 \left(\bar{c}^2\right)^{-\kappa^2} Z^2\right)^{1/\sigma},$$

$$\bar{N}^1 \left(\bar{c}^1\right)^{-\kappa^1} \left(c_D^1\right)^{1+\kappa^1} X^1 + \bar{N}^2 \left(\bar{c}^2\right)^{-\kappa^2} \left(c_D^2\right)^{1+\kappa^2} X^2 = 1.$$

Then, we can solve c_D^1 and c_D^2 from the following two equations:

$$\bar{N}^1 \left(\bar{c}^1\right)^{-\kappa^1} \left(c_D^1\right)^{1+\kappa^1} X^1 + \bar{N}_2 \left(\bar{c}^2\right)^{-\kappa^2} X^2 \left(\frac{\gamma_2^\sigma \bar{N}^1 \left(\bar{c}^1\right)^{-\kappa^1} Z^1}{\gamma_1^\sigma \bar{N}^2 \left(\bar{c}^2\right)^{-\kappa^2} Z^2}\right)^{\frac{\kappa^2+1}{\kappa^1+\sigma}} \left(c_D^1\right)^{\frac{\left(\kappa^1+\sigma\right)\left(\kappa^2+1\right)}{\kappa^2+\sigma}} = 1,$$

$$\bar{N}^1 \left(\bar{c}^1\right)^{-\kappa^1} X^1 \left(\frac{\gamma_1^\sigma \bar{N}^2 \left(\bar{c}^2\right)^{-\kappa^2} Z^2}{\gamma_2^\sigma \bar{N}^1 \left(\bar{c}^1\right)^{-\kappa^1} Z^1}\right)^{\frac{\kappa^1+1}{\kappa^1+\sigma}} \left(c_D^2\right)^{\frac{\left(\kappa^2+\sigma\right)\left(\kappa^1+1\right)}{\kappa^1+\sigma}} + \bar{N}^2 \left(\bar{c}^2\right)^{-\kappa^2} \left(c_D^2\right)^{1+\kappa^2} X^2 = 1.$$

Based on two equations, we know $\frac{\partial c_{1,D}}{\partial \bar{N}_1} < 0$. It is immediate that if ,

$$\frac{\partial c_{2,D}}{\partial \bar{N}_1} > 0, \frac{\partial L_2}{\partial \bar{N}_1} > 0, \frac{\partial L_1}{\partial \bar{N}_1} < 0, \quad \text{if } \sigma < 1,$$

$$\frac{\partial c_{2,D}}{\partial \bar{N}_1} < 0, \frac{\partial L_2}{\partial \bar{N}_1} < 0, \frac{\partial L_1}{\partial \bar{N}_1} > 0, \quad \text{if } \sigma > 1,$$

$$\frac{\partial c_{2,D}}{\partial \bar{N}_1} = 0, \frac{\partial L_2}{\partial \bar{N}_1} = 0, \frac{\partial L_1}{\partial \bar{N}_1} = 0, \quad \text{if } \sigma = 1.$$