Preferential Credit Policy with Sectoral Markup Heterogeneity *

Kaiji Chen[†] Yuxuan Huang[‡] Xuewen Liu [§] Zhikun Lu[¶]
Yong Wang[‖]

May 10, 2024

Abstract

Many emerging economies adopt preferential credit policies towards selected sectors. This paper provides an analysis of the economic rationale behind the preferential credit policies in the presence of market imperfections. Using firm-level data, we first empirically establish that sectors with higher markups are also those enjoying preferential credit subsidy. Disciplined by these empirical findings, we develop a multi-sector model featuring sectoral markup heterogeneity and endogenous firm entry and exit. We show that sector-specific preferential credit policies can be used to improve aggregate productive efficiency by reducing sectoral markup dispersion, despite the inefficiency introduced by allowing for less productive firms to enter and survive without exiting. We quantify the effect of preferential credit policy on aggregate TFP through the two opposite channels.

*We thank Joel David, Menglu Jiang, Pete Klenow, Ernest Liu, Yongseok Shin, Kjetil Storesletten, Shang-Jin Wei, Daniel Yi Xu, Mu-Jeung Yang, Xiaodong Zhu and participants of 2023 CES North American Conference, 1st annual Conference on China and the Global Economy, and 2023 International Conference on New Structural Economy for helpful discussions. The views expressed herein are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Atlanta.

[†]Emory University and Federal Reserve Bank of Atlanta

[‡]Institute of New Structural Economics, Peking University

[§]Department of Finance, HKU Business School, The University of Hong Kong

[¶]Goizueta Business School, Emory University

[‖]Institute of New Structural Economics, Peking University

1 Introduction

Many emerging and advanced economies adopt preferential credit policies as a kind of industrial policy towards selected sectors and firms. At first glance, the wide adoption of such policies might be puzzling from the standard view on resource misallocation, since they may exacerbate credit misallocation across sectors and entail efficiency loss by allowing inefficient firms to enter and stay in the economy. What are the economic rationales behind preferential credit policies? What are the impacts of such policy interventions on aggregate productive efficiency? Answers to these questions not only help to shed light on the role of such policies in emerging economies, but also help to understand to what extent policymakers should intervene to support specific industries.

This paper attempts to address these questions using China as a laboratory. China has arguably spent much more in helping favored industries with cheap loans than other major economies. According to Dipippo, Mazzocco and Kennedy (2022), China's industrial policy spending amounted to at least 1.73 percent of GDP in 2019, with below-market-rate bank credit being the largest component (0.52 percent of GDP), much larger than the scale of Japan (0.22 percent of GDP), the second-largest spender. In addition, China's preferential credit policy covers various industries from photovoltaics and LCD panels in the early 2000s to electronic vehicles more recently.

These policies have helped to increase the competitiveness of specific sectors, although those subsidized sectors often suffer from inefficient entry and exit. A prominent example is China's electric vehicle (EV) industry. The Chinese government designated electric vehicle (EVs) as a "strategic emerging industry" in 2009 and began flooding the sector with subsidies. The sector has met the government's targets after a decade. China's EV sales accounted for nearly 60% of global EV purchases. In 2023, China surpassed Japan as the world's top exporter, with China's BYD ousting Tesla as the world's top seller of EVs in 2023Q4.[1] On the other hand, the EV industry has long been troubled with low profit margins due to inefficient entry and exit. In 2014, for example, more than 80,000 companies registered in China to enter the EV sector, more than doubling the previous year's number of new registrants. As a result, despite strong sales growth, many domestic EV makers, including those top sellers, have long experienced profit losses and have relied on government subsidies to survive.[2]

[1]Another example is the LCD panel industry. In 2021 China took half of the world's market share in TFT-LCD panel production. In contrast, at the beginning of the 21st century, China relied entirely on imports of TFT-LCD panels for LCD TV production.

[2]For example, in 2022, three of the top ten EV manufacturers, Nio, Xpeng, and Li Auto, experienced profit losses of 14.4 billion yuan, 9.14 billion yuan, and 2.01 billion yuan, respectively. In 2020, Nio received a nearly $1 billion bailout from state-backed investors.

In this paper, we study how China's sector-specific preferential credit policies affect its aggregate productive efficiencies from both empirical and theoretical perspectives. We find that empirically, sectors with higher markups are those that enjoy preferential credit subsidies, which turn out to also have higher zombie firm ratios. Disciplined by this empirical finding, we develop a quantitative theory with sectoral markup heterogeneity and endogenous firm entry and exit. We show in such a framework that the preferential credit policy may be used as a second-best policy to alleviate the resource misallocation due to sectoral markup heterogeneity. The adoption of such policy, however, faces a trade-off between the efficiency gain by reducing the misallocation due to markup dispersion and the efficiency loss due to inefficient entry and exit.

In our empirical exercise, we establish the causal relationship between markup and credit subsidy, and the relationship between credit subsidy and the zombie ratio. Specifically, we first estimate firm-level markups and credit subsidies using firm-level data from the Annual Survey of Industrial Firms ("ASIF" henceforth). We then estimate the elasticity of credit subsidy to markup. To address the concern of reverse causality, we construct Bartik instruments using provincial-sector level markups. The estimated coefficient suggests a significant positive relationship between credit subsidy and markup. We further estimate the causal relationship between credit subsidy and zombie ratio at the sectoral level using the official lending rate as an instrument. We find robust evidence that higher credit subsidy leads to higher zombie ratio at the sectoral level.

To explain our empirical findings, we develop a multi-sector model in which preferential credit subsidies are positively related/linked to sectoral-level markups. Our model incorporates two key ingredients. The first key ingredient is sectoral-level markup heterogeneity, which generates intersectoral capital misallocation. Without credit subsidies, industries with above-average markups, i.e., those enjoying higher profit margins, tend to under-produce while industries with below-average markup tend to overproduce relative to the first-best case. This gives an incentive for well-intended governments to subsidize these industries with higher markups to equalize the marginal product of capital across industries. The second key ingredient is endogenous firm entry and exit for each industry. An industry's TFP level hence depends endogenously on the cut-off value of productivity for its firms to enter and exit, which in turn depends on the credit subsidy that the industry receives.

We show in such a framework that the industrial policy with heterogeneity in credit subsidy across industries simultaneously affects the dispersion of marginal revenue product of capital and the distribution of TFP across industries. Credit subsidy to industries with higher markups, on the one hand, reduces the dispersion of marginal revenue product of capital across industries, thus, improving allocative efficiency. On the other hand, such credit subsidy undermines the sectoral production efficiency by lowering the productivity threshold for a firm to enter and exit. In other words, the industrial policy with credit subsidy towards sectors with higher markups gives rise to a tradeoff between cross-sector capital misallocation and within-sector inefficient entry and exit.

To quantify the effects of preferential credit subsidy, we parameterize the credit subsidy scheme as a function of sectoral markup. We decompose the overall effect of preferential credit subsidy into two offsetting effects. First, an increase in the elasticity of credit subsidy to markup increases aggregate TFP by reducing the dispersion of marginal revenue product of capital due to markup dispersion. Second, it reduces aggregate TFP by increasing inefficiency in entry and exit. By calibrating our economy to China's firm-level data, we find that China's preferential credit subsidy policy leads to a significant increase in aggregate TFP relative to the laissez-faire economy, depsite that, compared with the second best policy, it over-subsidizes those sectors with high markups.

Our paper contributes to the growing literature on the role of industrial policy on emerging economies (e.g., Liu (2019) and Kim, Lee and Shin (2021)). The seminal work of Liu (2019) shows that when the downstream sectors suffer from market imperfections (e.g. due to financial constraint), the distortions pass through the upstream sectors by reducing the demand for intermediate inputs below the first-best level. This provides justification for the industrial policy in Korea and China to subsidize the upstream sectors as a second-best policy. Our paper complements Liu (2019) along several main dimensions. First, rather than focusing on distortions on the demand of the intermediate inputs due to frictions in the downstream sectors, we highlight the distortion on the supply of intermediate inputs as an outcome of sectoral markups to justify the preferential credit policy. Second, the market imperfections in Liu (2019) are generic and modeled as reduced-form wedges, while both our empirical evidence and theory highlight sectoral markup heterogeneity as a key friction for well-intended governments to intervene. Third, we examine the trade-off of such policy intervention in the context of firm dynamics and capital misallocation, an issue silent in Liu (2019). Our paper is also related Kim, Lee and Shin (2021), which shed lights on the trade-off of industrial policy. Using plant-level output and productivity data, Kim, Lee and Shin (2021) find that Korean government's policy of promoting heavy and chemical industries in the 1970s caused output and input use of targeted industries/regions to grow significantly faster than those of non-targeted ones. However, total factor productivity did not increase because the misallocation of resources across plants within targeted industries/regions got significantly worse. In our paper, the trade-off of industrial policy shows up as an exacebation of sectoral TFP due to inefficient entry and exit.

Our paper also contributes to the extensive literature on resource misallocation, especially from the perspective of China.[3] Using a structural model with reduced-form wedges, the seminal work of Hsieh and Klenow (2009) find large potential aggregate TFP gains from eliminating misallocation.[4] More recent work seek to quantify the role of financial frictions or policy distortions in aggregate TFP by discussing resource misallocation at the intensive and extensive margins. Using China's plant-level data, Midrigan and Xu (2014) argue that financial frictions do not generate sizable losses in aggregate TFP from misallocation, but from inefficient technological adoptions and barriers to entry. Wu (2018) estimates that policy distortions caused a large portion of capital misallocation, which accounted for the majority of the aggregate TFP loss. David and Venkateswaran (2019) shows that size-dependent policies or financial imperfections might be important sources of capital misallocation in China. Most papers in this literature assume different sources of distortions, either market imperfections or policy distortions, are independent from each other.[5] Accordingly, the first-best solution would be a completely removal of policy distortions or undo financial frictions created by preferential credit policy. By contrast, we argue that in the presence of market imperfections, a common feature in emerging economies, industrial policy, though distortive in nature, could be used to partially offset resource misallocation caused by market imperfections.

Finally, our paper relates to the recent literature on markup dispersion as a source of resource misallocation. Edmond, Midrigan and Xu (2015) shows that by exposing producers to greater competition, international trade may reduce markup dispersion thereby reducing misallocation and increasing aggregate productivity. Edmond, Midrigan and Xu (2023) study the welfare costs of markups in a dynamic model with heterogeneous firms and endogenously variable markups. They show that efficient allocation can be implemented by a specific nonlinear schedule of direct subsidies with two components, a uniform component that subsidizes all firms and that can be used to eliminate the aggregate markup, and a size-dependent component that used to eliminate misallocation. Our paper complements to the above studies by focusing on preferential credit subsidies as a second best policy to eliminate productivity dispersion caused by markup heterogeneity.

The rest of the paper is structured as follows. Section 2 establishes the empirical linkage between markup, credit subsidy and zombie ratio. Section 3 presents the model. Section 4 analyzes the role of preferential credit subsidy for aggregate TFP. Section 5 concludes.

---

[3]See Restuccia and Rogerson (2017) for an extensive review of this literature.

[4]Yang (2021), by extending the static framework of Hsieh and Klenow (2009) to a dynamic setting with endogenous occupation choices, shows that with endogenous selection, microfrictions can induce extensive margin misallocation among firms, as the composition of firms is shifted toward unproductive active firms.

[5]The literature has long argued that idiosyncratic distortions may well be correlated with the micro-level physical productivity, though. See, for example, Guner, Ventura and Xu (2008) and Restuccia and Rogerson (2008)

2 Empirical linkage between sectoral markup, credit subsidy, and zombie ratio

This section provides empirical evidence that disciplines our theoretical framework in the following sections. Specifically, we would establish the causal effect of sector-level markups on the average credit subsidy that the sector receives and the relationship between credit subsidy and zombie ratio at the sectoral level. Section 2.1 provides the details of constructing the data needed for our empirical investigation. In Section 2.2, we provide empirical linkage between markup and credit subsidy. Section 2.3 establish the linkage between credit subsidy and zombie ratio at the sectoral level.

2.1 Data Construction

The main dataset used for the markup calculation comes from the firm-level data (1998-2007) in the Annual Survey of Industrial Firms (ASIF) conducted by the National Bureau of Statistics of China. The dataset covers non-state-owned enterprises (non-SOEs) with annual sales of 5 million Chinese Yuan and all state-owned enterprises (SOEs). It is the most comprehensive firm-level dataset in China in terms of both firm and year coverage.

Following the literature, we compute the firm-level markups as the output elasticity of intermediate inputs divided by their cost share. The cost share of intermediate inputs can be directly calculated using the information in the dataset (expenditure on intermediate materials and sales revenue). To obtain the output elasticity of intermediate inputs, we estimate the production function at the four-digit sector level.[6] We then follow the procedure in De Loecker and Warzynski (2012) to obtain consistent estimates of the output elasticity. Appendix B provides the details of the estimation of firm-level markups.

After obtaining the firm-level markup for individual firm i at year t, denoted as $\mu_{i,t}$, we aggregate them to the four-digit sector level to construct the sector-level markups, denoted as $\mu_{s,t}$ for sector s, for all 421 manufacturing industries in our dataset. In particular, using firms' output shares $\frac{y_{i,t}}{\sum_{i\in s} y_{i,t}}$ as weights, we calculate the markup of sector s at year t as the weighted average of firms' markups within that sector:

$$\mu_{s,t} = \frac{\sum_{i\in s}\left(y_{i,t}\mu_{i,t}\right)}{\sum_{i\in s} y_{i,t}}, \tag{1}$$

We now compute the credit subsidy a firm receives. While the amount of credit subsidies firm receive are not directly reported in the ASIF data, it can be inferred from the difference

[6] See Appendix A for details.

between the minimum required interest rate based on the market rate and the effective interest rate. Following Caballero, Hoshi and Kashyap (2008), we compute the minimum required interest rate a firm i faces at year t as the weighted average of the short-term and midium-to-long-term benchmark lending rates.

$$R_t^i = rs_{t-1}\frac{BS_{t-1}^i}{B_{t-1}^i} + \left(\frac{1}{5}\sum_{j=1}^{5} rl_{t-j}\right)\frac{BL_{t-1}^i}{B_{t-1}^i},$$

where BS_{t-1}^i and BL_{t-1}^i denote short-term (less than one year) and midium-to-long-term bank loans, respectively, held by firm i at year $t-1$. rs_t and rl_t are the average short-term and long-term benchmark interest rates, respectively. $B_t^i = BS_{t-1}^i + BL_{t-1}^i$ is the total external liabilities of the firm at year t.[7] Following the calculation in Liu (2019), the effective lending interest rate, denoted as $EIR_t^i = \frac{IntExp_t^i}{B_t^i}$ is the ratio of the actual expenditure on interest reported by firms to their total external liabilities, conditioned on reporting positive interest payments.

With both minimum required interest rate and effective interest rate firm i faces at year t, the credit subsidy it receives at year t, denoted as τ_t^i, can then be computed as

$$\tau_t^i = 1 - EIR_t^i/R_t^i \tag{2}$$

We also construct the zombie ratio for a sector. To this end, we first use the so-called FN-CHK method, proposed by Caballero, Hoshi and Kashyap (2008) and Fukuda and Nakamura (2011), to identify whether a firm is a zombie. This method has been widely used in the literature to identify zombie firms.[8] Appendix C provides details of identification of zombie firms under this approach. We then construct zombie ratios as at the 4-digit sectoral level by dividing the total number of zombie firms by the total number of firms in the sector.

Table 1 reports the summary statistics for all variables used in our empirical regressions.

2.2 Linkage between markup and credit subsidy

In this section, we estimate the effects of sector-level markup on the credit subsidy a firm receives. Our hypothesis is as follows: the government tends to provide more credit subsidy to firms in sectors that enjoy higher profit margins to encourage firms in those sectors to expand their production scale. We conduct the following regression

[7] In the calculation of benchmark lending rate, we try various selections and combinations of official bank lending rates corresponding to different loan terms. The results remain robust.

[8] See, for example, Tan et al. (2017) for the application of this method for China.

$$\log\left(1-\tau_{s,t}^{i}\right)=\beta_0+\beta_1\mu_{s,t}+\beta_2X_{s,t}+\alpha_s+\alpha_t+\epsilon_{s,t}^{i}. \tag{3}$$

where $\tau_{s,t}^{i}$ is the subsidy rate for firm i in sector s at time t, $\mu_{s,t}$ is the 4-digit sectoral markup (logrithmically transformed), α_s, α_t and $\epsilon_{s,t}^{i}$ are the sector fixed effect, year fixed effect and the error term, respectively. Since credit subsidy at firm level might be affected by a firm's export status and ownership type, we include sector-level export share and SOE share as additional control variables $X_{s,t}^{i}$. The coefficient of interest is β_1, which captures the average effect of the markup on firms' credit subsidy rate at the sector level.

To alleviate the potential endogeneity problem, we employ the following estimation techniques: First, we keep the dependent variable at the firm level while the markup variable at the sector level to mitigate the reverse causality problem. Second, to reduce the endogeneity concerns from unobserved confounders, we employ the Bartik instrument approach in estimating equation (3). Following the standard decomposing approach, we can decompose the sector-level weighted average markups ($\mu_{s,t}$) into sector-province-level markups ($\mu_{s,p,t}$) and their corresponding sector-province-level output shares ($w_{s,p,t}$):

$$\mu_{s,t}\equiv\sum_{p}w_{s,p,t}\times\mu_{s,p,t}$$

The decomposition above is an accounting identity, which does not require specific assumptions. Based on this identity, the Bartik instrument is available as long as the individual components are observable. The firm-level dataset provides the possibility to calculate the components needed for instrumental variable construction. Specifically, the Bartik instrument ($z_{s,t}^{BI}$) can be constructed as the inner product of the sector-province-level output share in an initial year ($w_{s,p,0}$) and the province-level markups $\mu_{p,t}$ at year t

$$z_{s,t}^{BI}=\sum_{p}w_{s,p,0}\times\mu_{p,t}.$$

Following the convention in instrument construction (Goldsmith-Pinkham, Sorkin and Swift, 2020), we choose the share of a province p's output in sector s in the year 1998 as the predetermined provincial share in the Bartik instrument as it is the starting year in our data sample. By focusing on the differential impact of nationwide trends on distinct provinces, the instrument constructed above removes the variation in markups due to province-specific changes in sectoral structure or markup levels.

To estimate the regression in equation (3) using the constructed Bartik instrument, we implement a two-stage least squares (2SLS) estimation. The first-stage regression tests the relevance condition, which is typically satisfied for the Bartik instrument due to the way the instrumental variable is constructed.

$$\mu_{s,t} = \beta_0^F + \beta_1^F z_{s,t}^{BI} + \beta_2^F X_{s,t} + \alpha_s + \alpha_t + \epsilon_{s,t}^F$$

The 2SLS estimator can then be obtained by regressing $\log(1 - \tau_{s,t}^i)$ onto the first-stage fitted value $\hat{\mu}_{s,t}$ as shown below.

$$\log(1 - \tau_{s,t}^i) = \beta_0^S + \beta_1^S \hat{\mu}_{s,t} + \beta_2^S X_{s,t} + \alpha_s + \alpha_t + \epsilon_{s,t}^S. \tag{4}$$

Alternatively, we can estimate the following reduced-form regression, which gives a coefficient estimate quantifying the impact of the instrumental variable. The elasticity of credit subsidy to markups (β_1^S) can then be calculated using the reduced-form coefficient (β_1^R) divided by the estimated coefficient (β_1^F) in 2SLS estimation, i.e., $\beta_1^S = \beta_1^R / \beta_1^F$.

$$\log(1 - \tau_{s,t}^i) = \beta_0^R + \beta_1^R z_{s,t}^{BI} + \beta_2^R X_{s,t} + \alpha_t + \epsilon_{s,t}^R. \tag{5}$$

Table 2 presents the estimation result of β_1 for both OLS and IV regressions of equation (3). Column (1) shows that under OLS regression, the estimated β_1 is insignificant. On the other hand, the positive and statistically significant estimate ($\beta_1^F = 0.631$) suggests that the relevant condition is satisfied for the Bartik instrument. The second-stage coefficient estimate can be obtained through the calculation of dividing the reduced-form estimate by the first-stage estimate ($\beta_1^S = \beta_1^R / \beta_1^F = -0.408/0.631 = -0.648$), which is the same estimate given in column (4). The comparison of the IV estimate and the OLS estimate suggests that the endogeneity issue bias the OLS estimate, which is positive and insignificant. A possible reason for the insignificance under OLS regression is that while a higher markup or profit margin leads to higher credit subsidy, higher credit subsidy tends to encourage more firms to enter, which reduces the sectoral-level average markup.

In summary, our empirical finding establishes a positive relationship between markup and credit subsidy at the sector level.

2.3 Linkage between credit subsidy and zombie ratio

We now establish the empirical linkage between credit subsidy and zombie ratio at the sectoral level. Our hypothesis is that higher credit subsidy towards a sector tends to create more zombie firms at the sectoral level. This is because credit subsidy tends to benefit all firms in a sector, including those unproductive ones, thus encouraging those unproductive firms to stay, rather than exiting the market.

To examine this hypothesis, we run sector-level regressions to estimate the causal effects of credit subsidy on zombie ratio, controlling again the SOE share and export share. In reality, a higher zombie ratio may compel banks or government to provide a subsidized interest rate to keep firms surviving. To alleviate this potential reverse causality, we use the official lending rate as an instrument for τ in IV estimation. We use both the calculated credit subsidy (2) and its predicted value from the regression (4) as regressors.

Table 3 reports the estimated coefficients on various measures of credit subsidy. In Columns (1) and (2), both the estimated coefficients on the calculated τ are positive, suggesting that higher credit subsidy toward a sector would lead to higher zombie ratio. Columns (3) and (4) use the predicted τ as regressors. Again, the estimated coefficients on τ is positive and significant at 1 percent level. The point estimate in column (3) suggests that a one-percentage-point increase in credit subsidy τ would lead to a 1.885 percentage increase in zombie ratio.

In summary, our empirical findings suggest that higher credit subsidy would lead to higher zombie ratio at the sectoral level.

3 A Model of Preferential Credit Policy

We present a stylized model, which is disciplined by our empirical findings in the previous section on the relationship between markup and credit subsidy at the sector level. We use this model to explore the effects of preferential credit policy on aggregate TFP. Compared with Hsieh and Klenow (2009), our model has two distinct features: i) multiple sectors with heterogeneous markups of firms across sectors and ii) endogenous entry and exit of firms and thereby endogenous sectoral-level TFP. To focus on resource misallocation across sectors, we abstract from within-sector firm-specific distortions. Section 3.1 describes the model environment and Section 3.2 characterizes firms' entry and exit decision. Section 3.3 aggregates the economy and Section 3.4 analyzes the determinants of sector-level productivity.

3.1 Environment

We consider a partial equilibrium model, taking the supply side of capital as exogenous, to emphasize the decisions on firms' demand for capital. To be specific, the rental price of capital (interest rate) and the government credit subsidy rate are taken as exogenous.

Firms operate in multiple sectors. To highlight capital misallocation and keep the model as simple as possible, we assume that firms only use capital to produce output. All firms rent capital for production.

Sectors There are S sectors and each sector produces an intermediate composite good y_s. The final good Y is a CES aggregate of the S intermediate composite goods:

$$Y = \left(\sum_{s=1}^{S} y_s^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}, \tag{6}$$

where η is the elasticity of substitution among intermediate composite goods. In each sector s, there exists a continuum of firms and each firm i produces a differentiated product/variety $y_s(i)$. Differentiated products are transformed into the intermediate composite goods according to the following CES technology:

$$y_s = N_s^{\frac{1}{1-\varepsilon_s}} \left(\int_{i \in I_s} y_s(i)^{\frac{\varepsilon_s-1}{\varepsilon_s}} di \right)^{\frac{\varepsilon_s}{\varepsilon_s-1}}, \tag{7}$$

where i is the index for an active firm, I_s is the set of active firms in sector s, and N_s is the mass of active firms in sector s.[9] In particular, ε_s, which governs the markup, is sector-specific.

The markets of the final good Y and of the intermediate composite goods $\{y_s\}_{s=1}^{S}$ are perfectly competitive. As a result, the demand function for an intermediate composite good is given by

$$y_s = \left(\frac{p_s}{P} \right)^{-\eta} Y, \tag{8}$$

where p_s is the price of intermediate composite good s and $P = \left(\sum_{s=1}^{S} (p_s)^{1-\eta} \right)^{\frac{1}{1-\eta}}$ is the price of the final good. The demand curve for a differentiated product in sector s is given by

$$y_s(i) = \left(\frac{p_s(i)}{p_s} \right)^{-\varepsilon_s} \frac{y_s}{N_s}, \tag{9}$$

where $p_s(i)$ is the price of differentiated product i in sector s

$$p_s = \left(\frac{1}{N_s} \int_{i \in I_s} (p_s(i))^{1-\varepsilon_s} di \right)^{\frac{1}{1-\varepsilon_s}}. \tag{10}$$

Firms Firm i in sector s has access to the following AK production technology:

$$y_s(i) = a_s(i) k_s(i), \tag{11}$$

[9]The term $N_s^{\frac{1}{1-\varepsilon}}$ in (7) is to consider "no gain-to-variety". Under the alternative assumption that $N_s^{\frac{1}{1-\varepsilon}}$ is absent in (7), the conclusion of our model does not change.

where $a_s(i)$ is firm-specific productivity and $k_s(i)$ is variable capital input. For production to be operative, firms need to incur an overhead cost paid in terms of capital goods, denoted as $\bar{k}$. Firms rent capital at an exogenous interest rate R. The government provides sector-specific credit subsidies at rate τ_s, and thus, the effective rental rate is $R(1-\tau_s)$. It can be shown that the cost function for a firm with productivity $a_s(i)$ is given by

$$c(y_s(i); a_s(i), \tau_s) = mc_s\,(i)\; y_s(i) + c_{fs},$$

where the marginal/average cost $mc_s\,(i) = \frac{R(1-\tau_s)}{a_s(i)}$ and the fixed cost $c_{fs} = R(1-\tau_s)\bar{k}$.

Hence, firm i in sector s solves the following static profit maximization problem for each period:

$$\max_{p_s(i), y_s(i)} \; (p_s(i) - mc_s\,(i))y_s(i) - c_{fs} \quad \text{s.t.} \quad y_s(i) = \left(\frac{p_s(i)}{p_s}\right)^{-\varepsilon_s} \frac{y_s}{N_s}.$$

The optimal price is a constant markup multiplying the marginal cost:

$$p_s(i) = \mu_s\, mc_s\,(i) = \frac{\mu_s(1-\tau_s)R}{a_s(i)}, \tag{12}$$

where $\mu_s = \frac{\varepsilon_s}{\varepsilon_s - 1}$ is the sectoral markup over the marginal cost. The firm's profit is given by

$$\pi_s(a_s(i), \tau_s) = \Lambda_s \cdot \left(\frac{a_s(i)}{R(1-\tau_s)}\right)^{\varepsilon_s - 1} - c_{fs}, \tag{13}$$

where $\Lambda_s \equiv (\mu_s - 1)\,\mu_s^{-\varepsilon_s} p_s^{\varepsilon_s} \frac{y_s}{N_s}$.

3.2 Entry and Exit

We now consider a firm's entry and exit decision. Such decisions endogenize two variables for any sector s: the productivity cutoff $\underline{a}_s$ above which a firm will produce and otherwise not, and the total mass of active firms N_s. To reduce notational clutter without causing confusion, for the remaining of Section 3.2, we drop the sector index subscript s in notations; e.g., a_i means $a_s(i)$ and τ means τ_s.

To introduce endogenous exit, we assume for simplicity that a firm's productivity may shift across two states, i.e, normal state and low state. A firm i always starts with the normal state, in which the productivity level is a_i drawn from an initial productivity distribution upon entry. In each period, with probability $1-q$, a firm switches to the low state, in which the productivity becomes ρa_i with $\rho < 1$. The low state is an absorbing state in the sense that once a firm is in such a state, it will stay there forever. A firm will endogenously choose to exit if the value of running the business falls below zero. Furthermore, active firms also face an exogenous "death" shock with probability $1 - \phi$.

Firm Exit A firm will be active if and only if $\pi(a_i, \tau) \geq 0$, which, for any given credit subsidy τ, determines a productivity cutoff $\underline{a}$ that satisfies

$$\pi(\underline{a}(\tau), \tau) = 0. \tag{14}$$

Based on their productivity levels upon entry (i.e., in the normal state), we can divide firms within a sector into three types:

1. *Firms with $a_i < \underline{a}(\tau)$ in the normal state* Those firms will not produce even in the normal state and its value function in the normal state is

$$v^N(a;\tau) = 0.$$

2. *Firms with $a_i \in \left[\underline{a}(\tau), \frac{\underline{a}(\tau)}{\rho}\right)$ in the normal state* Those firms will produce in the normal state but will choose to exit after entering the low state. The value function of such a firm in the low state is $v^L(\rho a;\tau) = 0$ and the value function in the normal state is $v^N(a;\tau) = \pi(a;\tau) + \beta\phi[q\, v^N(a;\tau) + (1-q)v^L(\rho a;\tau)]$, which implies that

$$v^N(a;\tau) = \frac{\pi(a;\tau)}{1 - \beta\phi q}.$$

3. *Firms with $a_i \geq \frac{\underline{a}(\tau)}{\rho}$ in the normal state* Those firms will produce in the normal state and will choose not to exit even after they enter the low state. The value function of such a firm in the low state is $v^L(\rho a;\tau) = \pi(\rho a;\tau) + \beta\phi v^L(\rho a;\tau)$, that is, $v^L(\rho a;\tau) = \frac{\pi(\rho a;\tau)}{1-\beta\phi}$. The value function in the normal state is $v^N(a;\tau) = \pi(a;\tau) + \beta\phi[q\, v^N(a;\tau) + (1-q)v^L(\rho a;\tau)]$, which implies that

$$v^N(a;\tau) = \frac{\pi(a;\tau)}{1 - \beta\phi q} + \frac{\beta\phi(1-q)\frac{\pi(\rho a;\tau)}{1-\beta\phi}}{1 - \beta\phi q}.$$

Firm Entry To enter the market, a firm has to pay an entry cost c_e in terms of final goods. After paying the entry cost, the firm draws a productivity from a distribution $G(a)$ and starts to produce next period. The free entry condition states

$$\beta \int v^N(a;\tau) dG(a) \leq c_e. \tag{15}$$

In equilibrium, the free entry condition (15) holds as equality.

Without loss of generality, we assume that $G(a)$ is Pareto with tail parameter γ, that is, $G(a) = 1 - a^{-\gamma}$. Accordingly, a firm's expected value function in the normal state can be expressed as

$$\mathbb{E}v^N(a;\tau) = \chi \frac{c_f}{\underline{a}^\gamma}$$

where $\chi \equiv \frac{\rho(\varepsilon-1)(1-\beta\phi)-\beta(1-q)\phi\rho^{\gamma+1}(\gamma-\varepsilon+1)+\beta\gamma(1-q)\phi\rho^{\gamma+\varepsilon}}{\rho(1-\beta\phi)(\gamma-\varepsilon+1)(1-\beta q\phi)}$ is determined by exogenous model parameters, and we have used the result $\pi(a,\tau) = c_f\left[\left(\frac{a}{\underline{a}}\right)^{\varepsilon-1} - 1\right]$ by $\Lambda \cdot \left(\frac{\underline{a}}{R(1-\tau)}\right)^{\varepsilon-1} = c_f$ based on (14). Then, (15) implies $\beta\chi\frac{c_f}{\underline{a}^\gamma} = c_e$, that is,

$$\underline{a} = \left(\frac{\beta\chi c_f}{c_e}\right)^{\frac{1}{\gamma}} = \left(\frac{\beta\chi R(1-\tau)\bar{k}}{c_e}\right)^{\frac{1}{\gamma}}, \tag{16}$$

Note that $\underline{a}$ in (16) need to be equal to $\underline{a}(\tau)$ defined in (14) in equilibrium. Equation (16) shows that the level of cutoff productivity decreases with the credit subsidy rate τ: a higher credit subsidy would encourage firms with lower productivity to remain in production, rather than exiting compared with the case without credit subsidy.

Aggregate Firm Dynamics There are two groups of active firms in sector s: 1) those in the normal state and 2) those in the low state. Recall that active firms' productivity must satisfy $a_i \geq \underline{a}(\tau)$. Suppose the productivity of those in the normal state follows the Pareto distribution $\tilde{G}(a) = 1 - (a/\underline{a})^{-\gamma}$ in the support $a \in [\underline{a}, +\infty)$. Then, among them, those with productivity $a \in [\underline{a}/\rho, +\infty)$ will still stay in the sector after entering the low state; when this happens, their productivity distribution becomes $a' = a\rho \sim \tilde{G}(a') = 1 - (a'/\underline{a})^{-\gamma}$ in the support $a' \in [\underline{a}, +\infty)$. That is, after the negative productivity shock, those who enter the low state and stay have the same distribution in productivity as those in the normal state.

We focus on the stationary equilibrium, in which for any given sector the mass of firms in the normal state and the mass of firms in the low state remain constant over time. Let N_s^n be the mass of firms in the normal state, N_s^l be that in the low state, and $N_s = N_s^n + N_s^l$ be the total mass of active firms in sector s. We can derive the total mass of active firms in sector s by plugging Equation (16) into (14):

$$N_s = (\mu_s - 1)\,\mu_s^{-\varepsilon_s} p_s^{\varepsilon_s}\,(R(1-\tau_s))^{1-\varepsilon_s} \cdot \left(\frac{\beta\chi_s}{c_e}\right)^{\frac{\varepsilon_s-1}{\gamma}} c_{fs}^{\frac{\varepsilon_s-1-\gamma}{\gamma}} y_s, \tag{17}$$

which implies that higher credit subsidy leads to a larger number of active firms in that sector.

In stationary equilibrium, the inflow-outflow balance of low-state firms implies that

$$\underbrace{N_s^n \phi(1-q)\frac{1-G(\underline{a}_s/\rho)}{1-G(\underline{a}_s)}}_{\text{transformed from ``normal'' firms}} = \underbrace{N_s^l(1-\phi)}_{\text{exogenous exit}}. \tag{18}$$

Combining (17), (18), and $N_s = N_s^n + N_s^l$, we can pin down N_s, N_s^n, and N_s^l. Also, in the stationary equilibrium, the mass of new entrants equals the mass of "normal" firms that exit the market or become "low" firms, by which we can determine the mass of firms who participate in drawing the productivity lottery, N_s^e:

$$N_s^e\left(1-G\left(\underline{a}_s\right)\right) = \underbrace{N_s^n(1-\phi)}_{\text{exogenous exit}} + \underbrace{N_s^n\phi(1-q)}_{\text{endogenous exit or becoming ``low'' firms}},$$

where $N_s^e\left(1-G\left(\underline{a}_s\right)\right)$ is the mass of new entrants with productivity above the cutoff productivity, i.e., those entrants that are operative. Recall that some firms pay entry cost but choose not to operate if its productivity draw is sufficiently low. Note that $N_s^n\phi(1-q) = N_s^n\phi(1-q)\frac{G(\underline{a}_s/\rho)-G(\underline{a}_s)}{1-G(\underline{a}_s)} + N_s^l(1-\phi)$, including those firms (with productivity $[\underline{a}_s, \underline{a}_s/\rho]$ upon entry) that have entered the low state, and those "low" firms that exit exogenously.

3.3 Aggregation

Plugging Equation (12) into(10) and rearranging, we derive the sectoral price

$$p_s = \mu_s R(1-\tau_s)\left(\int_{a=\underline{a}_s}^{\infty} a^{\varepsilon_s-1} d\tilde{G}(a)\right)^{\frac{1}{1-\varepsilon_s}} = \mu_s \frac{R(1-\tau_s)}{\bar{a}_s}, \tag{19}$$

where $\tilde{G}(a) = \frac{G(a)-G(\underline{a})}{1-G(\underline{a})}$ is the productivity distribution of incumbent firms, and

$$\bar{a}_s \equiv \left(\int_{a=\underline{a}_s}^{\infty} a^{\varepsilon_s-1} d\tilde{G}(a)\right)^{\frac{1}{\varepsilon_s-1}} \tag{20}$$

is the weighted power mean productivity of a sector s. In the next section, we show that $\bar{a}_s$ is equal to the TFP of sector s.

The aggregate profit of firms in sector s is given by

$$\Pi_s = \int_{i\in I_s} \pi_s(a_s(i), \tau_s) di = N_s \pi_s(\bar{a}_s, \tau_s),$$

where the second equality is obtained by plugging (13).[10] The above equation suggests that the profit of individual firms (Π_s/N_s) is positively affected by the TFP of the sector. Thus, an increase of the credit subsidy to a specific sector tends to decrease the average profitability of the sector.

The aggregate demand for variable capital inputs in sector s is

$$K_s = \int_{i\in I_s} k_s(i)\, di,$$

For a given interest rate R and subsidy scheme $\tau = \{\tau_s\}_{s=1}^{S}$, the aggregate demand for capital in the economy is

$$K = \sum_{s=1}^{S} \left(K_s + \bar{k}N_s\right).$$

Comment on Equilibrium Similar to Hsieh and Klenow (2009), our model adopts a partial equilibrium framework with a focus on production. We specifically examine capital (mis)allocation on the supply side (i.e., firms) whose efficiency is measured by aggregate TFP, abstracting away from the demand side (i.e., households). For a given aggregate demand Y, we can determine how much output each sector produces and how much input (in terms of capital goods) each requires. We can then calculate the TFP by Y/K. The aggregate production function features constant returns to scale and behaves like AK technology, where A is endogenously determined by both between-sector and within-sector resource allocation efficiency. As we will see in Section 4, capital goods are inefficiently allocated due to the monopoly markup in the product market and the credit subsidy in the factor market; however, a proper credit policy can mitigate TFP loss in the presence of sectoral markup heterogeneity.

3.4 Sector-level Productivity

The physical productivity of a sector can be defined as

$$\mathrm{TFP}_s \equiv \frac{y_s}{K_s},$$

Lemma 1 *The physical productivity of a sector equals to its average productivity*

$$\mathrm{TFP}_s = \bar{a}_s. \tag{21}$$

[10] See Appendix D for the detailed derivation.

Proof: See Appendix D.

Plugging $\tilde{G}(a) = 1 - (a/\underline{a})^{-\gamma}$ into equation (20) and rearranging yields

$$\bar{a}_s = \left(\frac{\gamma}{\gamma - \varepsilon_s + 1}\right)^{\frac{1}{\varepsilon_s - 1}} \underline{a}_s. \tag{22}$$

Equation (22) implies that the physical productivity of a sector s is proportional to its cutoff productivity $\underline{a}_s$. Together with equation (16), it implies that an increase in credit subsidy reduces a sector's TFP.

We now derive the revenue-productivity for a sector. The marginal revenue product of capital (MRPK) for firm i in sector s is given by

$$\text{MRPK}_s(i) \equiv p_s(i)a_s(i) = \mu_s(1 - \tau_s)R,$$

where the second equality is obtained by plugging (12). Note that MRPK is homogeneous across all firms within a sector. Recalling (19), the price index of sector s is given by

$$p_s = \frac{\mu_s(1 - \tau_s)R}{\bar{a}_s},$$

which implies that the revenue productivity of a sector s is given by

$$\text{TFPR}_s \equiv p_s\bar{a}_s = \mu_s(1 - \tau_s)R.$$

Note that since we abstract from labor and the production function is linear in capital, it follows that MRPK_s=TFPR_s for all s.

Overall, for a given subsidy scheme $\tau = \{\tau_s\}_{s=1}^S$, we can find out the distribution of the marginal revenue product of capital, $\{\text{MRPK}_s\}_{s=1}^S$ (which is also $\{\text{TFPR}_s\}_{s=1}^S$), and the sectoral-level TFP, $\{\bar{a}_s\}_{s=1}^S$.

4 Implications for Aggregate TFP

In this section, we analyze the determinants of aggregate TFP. Section 4.1 illustrates aggregate TFP as a function of preferential credit subsidy in a two-sector economy with different markups. Section 4.2 then parameterizes preferential credit subsidy as a loglinear function of sectoral markups in our baseline economy with mutiple sectors and provides a decomposition of aggregate TFP. Section 4.3 calibrates our baseline economy to Chinese economy and

Section 4.4 provides a quantitative analysis of optimal preferential credit policy in such a framework.

4.1 An Illustrative Example

In this section, we illustrate with a numerical example of a two-sector economy how the level of credit subsidy towards the sector with higher markups influences aggregate TFP. Specifically, Sector 1 is monopolistically competitive and Sector 2 is perfect competitive. For Sector 1, therefore, the markup $\mu_1 = 1$, and for Sector 2, we set it as $\mu_2 = \frac{4}{3}$. We normalize the credit subsidy in Sector 1 to be zero ($\tau_1 = 0$) and study the optimal credit subsidy to Sector 2 (τ_2).

If the entry and exit is not allowed (in which case the sector-level productivity $\bar{a}_s$ is exogenous), the optimal subsidy policy τ_2 is set to counteract the markup effect; that is, the optimal subsidy policy is such that in equilibrium the marginal revenue product of capital, $\{\text{MRPK}_s\}_{s=1}^2$, is homogeneous across sectors. Hence, in this numerical example, the optimal policy is to set the subsidy ratio for the monopolistically competitive sector as $\tau_2 = \frac{1}{4}$. This way, the MRPK for the monopolistically competitive sector is given by $\mu_2 R(1-\tau_2) = R$, the MRPK of the perfectly competitive sector. However, with endogenous entry and exit, there is an extra cost associated with subsidy, i.e., lowering the average productivity of the subsidized sector. So the optimal subsidy should be lower under the case of endogenous entry and exit than under the case of no entry or exit.

To show the result formally, we define "net" aggregate TFP (total output over total capital stock net of overhead fixed costs) as

$$\text{TFP}^n \equiv \frac{Y}{K^n},$$

where $K^n = \sum_{s=1}^2 K_s$ is the total amount of capital excluding overhead fixed costs. Figure 1 shows the TFP gain as a function of the subsidy ratio for the monopolistically competitive sector, i.e., TFP gain (%) = $\log(\text{TFP}(\tau_2)) - \log(\text{TFP}(\tau_2 = 0))$ as a function of τ_2. For each panel, which corresponds to different papameter values of η, the elasticity of substitution across sectors, there exists an optimal subsidy ratio. For example, when $\eta = 0.5$, aggregate TFP is hump-shaped in our benchmark economy with endogenous entry and exit, with the optimal credit subsidy to sector 1 around 0.15, under which aggregate TFP peaks at 0.5%. By contrast, in the economy without firm entry and exit, the optimal credit subsidy to sector 1 equals to 0.25, much higher than that in the benchmark economy. This suggests that in the economy with endogenous entry and exit, credit subsidy generates the efficiency loss by allowing inefficient firm to remain in production. Thus, the optimal credit subsidy trades off the efficiency gain due to reduction in markup dispersion and the efficiency loss in the sector enjoying credit subsidy.

It is interesting to see how η, the elasticity of substitution across sectors, affects the gap in the optimal subsidy ratio. Under the case with endogenous entry and exit, the optimal credit subsidy is smaller under a smaller η.[11] Intuitively, the smaller η is, the smaller gains from equalizing TFPR, which can been see from a much lower aggregate TFP gain with the optimal credit subsidy under the case $\eta = 0.5$ relative to the case $\eta = 2$ (0.5% versus 2%). A smaller potential TFP gain causes the optimal credit subsidy rate to be lower to offset the efficiency loss due to endogenous firm entry and exit.

4.2 Aggregate TFP in the multi-sector economy

In this section, we generalize our analysis of the role of preferential credit subsidy on aggregate TFP in the multi-sector economy, which will be used to calibrate to the Chinese economy in the next section. We first analyze the impacts of credit subsidy on net aggregate TFP, followed by their impacts on aggregate TFP, which takes into account efficiency loss due to overhead fixed costs.

To begin with, we first characterize the aggregate TFP for a given markup distribution $\boldsymbol{\mu} = \{\mu_s\}_{s=1}^S$ and a given subsidy scheme $\boldsymbol{\tau} = \{\tau_s\}_{s=1}^S$.

Plugging (8) into (6), we can rewrite the final output as

$$Y = \left(\sum_{s=1}^{S} \left(\frac{p_s}{P} \right)^{1-\eta} \right)^{\frac{\eta}{\eta-1}} Y. \tag{23}$$

The total *net* demand for capital (net of fixed capital inputs) is given by

$$K^n = \sum_{s=1}^{S} K_s = \sum_{s=1}^{S} \frac{y_s}{\bar{a}_s} = \sum_{s=1}^{S} \left(\frac{p_s}{P} \right)^{-\eta} Y \frac{1}{\bar{a}_s}, \tag{24}$$

where the second equality is obtained using the sector-level aggregate production function $y_s = \bar{a}_s K_s$ and the third equality is obtained by plugging (8).

With (23) and (24), the (net) aggregate TFP can be measured as

$$\mathrm{TFP}^n \equiv \frac{Y}{K^n} = \frac{\left(\sum_{s=1}^{S} \left(\frac{p_s}{P} \right)^{1-\eta} \right)^{\frac{\eta}{\eta-1}} Y}{\sum_{s=1}^{S} \left(\frac{p_s}{P} \right)^{-\eta} Y \frac{1}{\bar{a}_s}} = \frac{\left(\sum_{s=1}^{S} p_s^{1-\eta} \right)^{\frac{\eta}{\eta-1}}}{\sum_{s=1}^{S} p_s^{-\eta} \frac{1}{\bar{a}_s}}.$$

[11] Without firm entry and exit, the optimal credit subsidy rate, 1/4, is the same across different values of the η.

Plugging (19) into the above equation, we have

$$\mathrm{TFP}^n = \frac{\left(\sum_{s=1}^{S} \left(\mu_s(1-\tau_s)\right)^{1-\eta} (\bar{a}_s)^{\eta-1}\right)^{\frac{\eta}{\eta-1}}}{\sum_{s=1}^{S} \left(\mu_s(1-\tau_s)\right)^{-\eta} (\bar{a}_s)^{\eta-1}}. \tag{25}$$

Note that in (25), $\bar{a}_s$ is a function of the sector-specific subsidy ratio τ_s.

To obtain a closed-form solution for aggregate TFP, we assume that the subsidy scheme follows a simple rule

$$\tau_s = \tau(\mu_s) = 1 - \left(\frac{\mu_s}{\overline{\mu}}\right)^{\frac{1}{\alpha}-1}, \tag{26}$$

where $\overline{\mu}$ is the average markup of the economy. In equation (26), the subsidy is non-decreasing in markups and the parameter $\alpha \geq 1$ controls the elasticity of credit subsidy to markups at the sectoral level. For the extreme case with $\alpha = 1$, we have $\mu_s(1-\tau_s) = \mu_s$, which corresponds to the case of no subsidy; for the other extreme case with $\alpha \to \infty$, we have $\mu_s(1-\tau_s) = \overline{\mu}$, which fully eliminates the MRPK dispersion across sectors. Figure 2 shows $\mu_s(1-\tau_s)/\overline{\mu}$ as a function of $\mu_s/\overline{\mu}$ under different α. The larger is α, the smaller is the slope of the function $\mu_s(1-\tau_s)$ with respect to μ_s. The case of $\alpha = 1$ corresponds to the 45 degree line. When α is as large as 100, the MRPK dispersion across sectors is basically eliminated. Plugging (26) into (25), we can obtain aggregate TFP as

$$\mathrm{TFP}^n = \frac{\left(\sum_{s=1}^{S} (\mu_s)^{\frac{1-\eta}{\alpha}} (\bar{a}_s)^{\eta-1}\right)^{\frac{\eta}{\eta-1}}}{\sum_{s=1}^{S} (\mu_s)^{-\frac{\eta}{\alpha}} (\bar{a}_s)^{\eta-1}}. \tag{27}$$

Note that in standard models such as Hsieh and Klenow (2009), the TFP of sector s, $\bar{a}_s$, is exogenous and uncorrelated with the markup μ_s. Suppose μ_s follows log-normal distribution, i.e., $\log(\mu_s) \sim N(\bar{\mu}, \sigma_\mu^2)$. Using the Central Limit Theorem and assuming $S \to \infty$, we obtain the following decomposition for aggregate TFP

$$\log(\mathrm{TFP}^n) = \frac{1}{\eta-1}\log\left(\sum_{s=1}^{S} (\bar{a}_s)^{\eta-1}\right) - \frac{\eta}{2\alpha^2}\sigma_\mu^2. \tag{28}$$

The first argument on the right-hand-side of (28) is the efficient level of TFP without capital misallocation. The second argument captures the capital misallocation acoss sectors due to markup dispersion, which is decreasing in α. That is, the subsidy scheme to counteract the dispersion of capital productivity due to markup heterogeneity improves allocative efficiency. For the extreme case with $\alpha = 1$ (no subsidy), we have $\log(\mathrm{TFP}^n) = \frac{1}{\eta-1}\log\left(\sum_{s=1}^{N} (\bar{a}_s)^{\eta-1}\right) - \frac{\eta}{2}\sigma_\mu^2$, similar to the result in Hsieh and Klenow (2009). For the other extreme case with

$\alpha \to \infty$ (fully eliminating TFPR dispersion), we have $\log(\text{TFP}^n) = \frac{1}{\eta-1}\log\left(\sum_{s=1}^{N}(\bar{a}_s)^{\eta-1}\right)$, which is the efficient level of TFP.

Different from Hsieh and Klenow (2009), in our model, $\bar{a}_s$ and μ_s are correlated due to the following two reasons. First, μ_s directly affects the level of TFP of a sector via endogenous entry and exit. Second, through our credit subsidy scheme, μ_s affects $\underline{a}_s$, the threshold productivity for a firm to be operative. According to Equation (16), $\underline{a}_s(\tau_s)$ is a decreasing function of τ_s. Thus, for some firms in the low state (i.e., those with $\rho a_s(i) \in [\underline{a}_s(\tau_s), \underline{a}_s(0))$ after shock), if there were no subsidies, they would have exited the market when hit by the negative productivity shock. We call those firms as zombie firms. This is an inefficiency on the exit side.[12]

To obtain the marginal impacts of the subsidy scheme τ_s on $\bar{a}_s$, we can rewrite $\bar{a}_s$ by a combination of (16) and (22)

$$\bar{a}_s = a_s(1-\tau_s)^{\frac{1}{\gamma}}, \tag{29}$$

where $a_s \equiv \left(\frac{\gamma}{1+\gamma-\varepsilon_s}\right)^{\frac{1}{\varepsilon_s-1}}\left(\frac{\beta\chi_s R\bar{k}}{c_e}\right)^{\frac{1}{\gamma}}$ is independent of τ_s. The term $1/\gamma$ captures the elasiticity of $\bar{a}_s$ with respect to $1-\tau_s$. Equation (29) implies that the magnitude of the marginal impact of τ_s depends negatively on γ.

Plugging equation (26) and (29) into equation (27) and rearranging, we can express net aggregate TFP as

$$\text{TFP}^n = \frac{\left(\sum_{s=1}^{S}(\mu_s)^{\frac{1-\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}(a_s)^{\eta-1}\right)^{\frac{\eta}{\eta-1}}}{\sum_{s=1}^{S}(\mu_s)^{-\frac{\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}(a_s)^{\eta-1}}. \tag{30}$$

Comparing equation (30) with (27), we see that unless $\gamma = \infty$, aggregate TFP under endogenous entry and exit is not equal to its counterpart with exogenous sector-level TFP. We have the following proposition

Proposition 1 *Under the assumption that μ_s and of a_s follows a joint log-normal distribution and $S \to \infty$, net aggregate TFP can be decomposed as*

$$\begin{aligned}\log(\text{TFP}^n) =& \frac{1}{\eta-1}\log\left(\sum_{s=1}^{N}(a_s)^{\eta-1}\right) - \frac{\eta}{2\alpha^2}\sigma_\mu^2 - \bar{\mu}\left(1-\frac{1}{\alpha}\right)\frac{1}{\gamma} \\ &- \left(1-\frac{1}{\alpha}\right)\frac{\eta-1}{\gamma}Cov\left(\log(\mu_s),\log(a_s)\right) + \left(1-\frac{1}{\alpha}\right)^2\left(\frac{\eta-1}{\gamma^2}\right)\frac{\sigma_\mu^2}{2}.\end{aligned} \tag{31}$$

[12] There is also an inefficiency on the entry side. Specifically, for some firms in the normal regime (i.e., those with $a_s(i) \in [\underline{a}_s(\tau_s), \underline{a}_s(0))$ before shock), they would not have entered the market if there were no subsidies.

Proof: See Appendix D.

Comparing with the right-hand-side of (28), equation (31) contains three new terms. The first new term, $\bar{\mu}\left(1-\frac{1}{\alpha}\right)\frac{1}{\gamma}$, captures the *average* effect of markup on efficiency loss via its impact on credit subsidy. Note that according to (26), $\left(1-\frac{1}{\alpha}\right)$ is the elasticity of $1-\tau_s$ with respect to the markup μ_s, while according to (29), $\frac{1}{\gamma}$ is the elasticity of sector-specific productivity $\bar{a}_s$ with respect to $1-\tau_s$. Hence, the coefficient $\left(1-\frac{1}{\alpha}\right)\frac{1}{\gamma}$ for $\bar{\mu}$ in equation (31) can be interpreted as the elasticity of sector-specific productivity $\bar{a}_s$ to markup μ_s via its impact on credit subsidy. Hence, with the average markup level, $\bar{\mu}$, as the multiplier, this term reduces aggregate TFP due to inefficient entry and exit.

The second new term, $\left(1-\frac{1}{\alpha}\right)\frac{\eta-1}{\gamma}Cov\left(\log(\mu_s),\log(a_s)\right)$, moderates the strength of the first new term; it only appears when $Cov\left(\log(\mu_s),\log(a_s)\right)\neq 0$. Note that according to equation (29), the impact of subsidy $1-\tau_s$ on sector-level TFP is proportional to a_s. Thus, for a more productive sector, credit subsidy τ leads to more efficiency loss due to inefficient entry and exit. When $Cov\left(\log(\mu_s),\log(a_s)\right)>0$, more productive sectors tend to have higher markup, thus receiving more subsidies according to equation (26). As a result, the overall negative impact of subsidy on aggregate TFP will be amplified under a positive $Cov\left(\log(\mu_s),\log(a_s)\right)$.[13] Notice that by the definition of a_s, $Cov\left(\log(\mu_s),\log(a_s)\right)$ is invariant with the credit subsidy scheme α.[14]

We measure net TFP loss as the gap between the log of net aggregate TFP and its efficient level:

$$\begin{aligned}
\text{Net TFP loss (\%)} &= \log(\text{TFP}^n)-\frac{1}{\eta-1}\log\left(\sum_{s=1}^{S}(a_s)^{\eta-1}\right)\\
&= \underbrace{\left[\frac{\eta}{\alpha^2}-\left(1-\frac{1}{\alpha}\right)^2\left(\frac{\eta-1}{\gamma^2}\right)\right]\frac{\sigma_\mu^2}{2}}_{\text{efficiency loss due to markup dispersion}}\\
&\quad \underbrace{+\bar{\mu}\left(1-\frac{1}{\alpha}\right)\frac{1}{\gamma}+\left(1-\frac{1}{\alpha}\right)\frac{\eta-1}{\gamma}Cov\left(\log(\mu_s),\log(a_s)\right)}_{\text{efficiency loss due to inefficient entry and exit}} \qquad (32)
\end{aligned}$$

As Equation (32) shows, net TFP loss originates from two sources, the first capturing the

[13] Additionally, it requires $\eta>1$. Consider a simple example where $\eta=2$, $N=2$, and $\log(\text{TFP})=\log(a_1+a_2)$. Suppose only one sector has market power and thus receives credit subsidy; when subsidy is imposed, the sector level TFP will drop by 10%. Let $a_2=2>a_1=1$. Then, we have $\log(0.9a_1+a_2)=1.065>\log(a_1+0.9a_2)=1.030$. We can see that the TFP is lower when the productive sector (i.e., sector 2) receives subsidy (say due to higher markup).

[14] The third new term $\left(1-\frac{1}{\alpha}\right)^2\left(\frac{\eta-1}{\gamma^2}\right)\frac{\sigma_\mu^2}{2}$ is a higher-order interaction term, which helps to increase the aggregate TFP.

efficiency loss due to markup dispersion and the second the TFP loss due to inefficient entry and exit. Note that as α increases, both terms in the efficiency loss due to markup dispersion become smaller, implying improvement in allocative efficiency. On the other hand, both terms in the efficiency loss due to inefficient entry and exit increases, since a higher elasticity of credit subsidy to markup causes larger TFP loss due to inefficient entry and exit, and the more so if the sector with higher markup is more productive. We call the first term in each of the two sources of efficiency loss as the main effect, and the second term the interactive effect.

Now we analyze the impacts of preferential credit subsidy on aggregate TFP, measured as total output over total capital input

$$\text{TFP} \equiv \frac{Y}{K} = \text{TFP}^n \times \frac{K^n}{K} \tag{33}$$

where

$$K = \sum_{s=1}^{S} \left(K_s + \bar{k} N_s \right). \tag{34}$$

is the total capital demand of the economy, including both the variable and fixed inputs.

Plugging (34) into (33) and rearranging, we can decompose aggregate TFP as

$$\log(\text{TFP}) = \log(\text{TFP}^n) - \underbrace{\log\left(1 + \frac{\sum_s N_s \bar{k}}{\sum_s K_s}\right)}_{\text{TFP loss due to overhead costs}},$$

where the second argument on the right-side of the above equation captures the additional TFP loss due to overlead costs. When all sectors have the same markups, we have

$$\frac{\sum_s N_s \bar{k}}{\sum_s K_s} = \frac{(\mu - 1)(\gamma - \epsilon + 1)}{\gamma}, \tag{35}$$

which is independent of $\bar{k}$, c_e, and α.[15] However, with sectoral heterogeneity, a higher subsidy causes a higher share of capital goods being used as overhead costs, which lowers TFP. Accordingly, we can measure total TFP loss as the sum of net TFP loss and TFP loss due to overhead costs:

$$\text{Total TFP loss } = \text{Net TFP loss} + \log\left(1 + \frac{\sum_s N_s \bar{k}}{\sum_s K_s}\right)$$

[15] See Appendix D for a derivation.

4.3 Calibration

We calibrate our model economy to match key data moments from the 2006 Annual Survey of Industrial Firms (ASIF). A period in our economy is one year and we set the discount factor β to be 0.96 and the gross interest rate $R = 1.04$. Regarding the elasticity of substitution across intermediate goods, although available estimates of η vary widely across studies, most of them are in the range [2,10] and a value around $\eta = 5$ is most frequently used in quantitative exercises (e.g., Epifani and Gancia, 2011). We therefore set $\eta = 5$ as a benchmark.

The level and dispersion of TFPR and its impacts on aggregate TFP depends on two key parameters. The first is α, which governs the sensitivity of credit subsidy to markup. To calibrate α, we would like to map the theoretical response of credit subsidy to sector-level markup into its empirical counterpart. Taking the logarithm of both sides of equation (26), we have

$$\log\left(1-\tau_s\right) = \left(\frac{1}{\alpha}-1\right)\log\left(\mu_s\right) + \left(1-\frac{1}{\alpha}\right)\log\overline{\mu} = \beta_0 + \beta_1 \log\left(\mu_s\right) \tag{36}$$

Comparing (36) with its empirical counterpart (3), we have $\alpha = \frac{1}{1+\beta_1}$. Since $\beta_1 = -0.648$ from our empirical estimation, we have $\alpha = 1/(1-0.648) = 2.84$.

A second key parameter is γ, the tail parameter of the productivity distribution. We calibrate γ to target the average revenue share of the top 5% firms in 4-digit ASIF industries. For 2006, the top 5% revenue share is 43.53%. In Appendix E, we show this target maps into a parameter value of $\gamma = 7.68$ for the Pareto distribution of firm-level productivity.

We use the following three moments to jointly calibrate $\{\rho, \phi, q\}$ by solving a nonlinear system of equations:

1. The ratio of entrant firms in the total numer of firms:

$$\frac{N^e(1-G(a))}{N} = \frac{(1-\phi)(1-q\phi)}{1-\phi+(1-q)\rho^\gamma\phi}$$

2. The sales share of exiting firms in total sales:

$$\begin{aligned}&(1-\phi)+\phi\frac{N^n(1-q)\dfrac{\int_{\underline{a}}^{\frac{\underline{a}}{\rho}}\Omega a^{\varepsilon-1}d\tilde{G}(s)}{\int_{\underline{a}}^{\infty}\Omega a^{\varepsilon-1}d\tilde{G}(s)}}{N}\\ =&(1-\phi)+\phi(1-q)\frac{(1-\phi)}{1-\phi+(1-q)\rho^\gamma\phi}\left(1-\rho^{1+\gamma-\epsilon}\right)\end{aligned}$$

   Under a larger ρ, more productive firms will exit, which increases the sales share of exiting firms.

3. The ratio of zombie firms in the total numer of firms:

$$\frac{N^l\left(\frac{G(\underline{a}^0)-G(\underline{a})}{1-G(a)}\right)}{N}=\frac{N^l}{N}\frac{s}{R}=\frac{(1-q)\rho^\gamma\phi}{(1-\phi+(1-q)\rho^\gamma\phi)}\frac{s}{R}$$

where $\underline{a}^0$ is the productivity cutoff without credit subsidy. For the year 2006, we have the share of entrant firms in total firms equal to 0.020; The sales share of exiting firms in total sales equal to 0.016; and the share of zombie firms in total firms as 0.230. Thus, we have $\rho = 0.953, \phi = 0.986, q = 0.761$.

The remaining undetermined parameters are c_e and $\bar{k}$. It turns out only the ratio $\frac{c_e}{\bar{k}}$ matters for the magnitude of TFP. We set $\frac{c_e}{\bar{k}} = 100$. Our quantitative results, however, are insensitive to the value of $\frac{c_e}{\bar{k}}$.

Note that the only source of sectoral heterogeneity is the markups $\{\mu_s\}_{s=1}^S$. To compute $Cov(\log \mu_s, \log a_s)$, we first draw $\log(\mu_s) \sim N(0.166, 0.085^2)$, the estimated distribution from our constructed sector-level markups. Since the model requires $\gamma - \varepsilon_s + 1 > 0$, we discard draws that fall beyond our model's acceptable range. Next, we compute a_s and then $Cov(\log \mu_s, \log a_s)$. Note that $a_s = \left(\frac{\gamma}{1+\gamma-\varepsilon_s}\right)^{\frac{1}{\varepsilon_s-1}}\left(\frac{\beta\chi_s R\bar{k})}{c_e}\right)^{\frac{1}{\gamma}}$ does not depend on the policy parameter α.

Table 4 summarizes the parameter values for our baseline economy.

4.4 Optimal preferential credit policy

This section explores the optimal preferential credit policy. We first explore the optimal credit policy that minimize the net TFP loss. Panel (a) of Figure 3 shows net TFP loss as a function of the policy parameter α. We see that total TFP loss is u-shaped in α. Without preferential credit subsidy, markup dispersion causes an aggregate TFP to be 1.806 percent below the efficient TFP level. As α increases, sectors with higher markup enjoy more credit subsidy, which encourage firms with higher markup to produce more, thus, reducing the TFPR dispersion. As a result, net TFP loss starts to decline and it reaches it lowest point, 1.322 percent at $\alpha = 2.09$. In other words, optimal credit subsidy would reduce the aggregate TFP loss due to markup dispersion by about one fourth (1.322/1.806-1).

When α goes beyond this optimal level, net TFP loss starts to increase. This suggests that preferential credit policy trades off efficiency loss due to markup dispersion against efficiency loss due to inefficient entry and exit. Such tradeoff is illustrated in Panel (b) of Figure 3, which decomposes net TFP loss into these two opposing forces. As we can see, as α increases from 1 to 8, efficiency loss due to markup dispersion reduces from over 1.8% to less than 0.1%. By contrast, efficiency loss due to inefficient entry and exist increases from 0 to more than 1.5%. Both Panel (c) and (d) show that the main efect is the dominating force driving the efficiency loss arising from these two competing forces.

What does our results imply for the preferential credit subsidy adopted by China? Note that in Section 2, our estimated subsidy policy parameter $\alpha = 2.84$. This implies that the current preferential credit subsidy policy raises aggregate TFP significantly compared with the case without credit subsidy, though it over-subsdizes sectors with higher markups relative to the second-best policy.

Figure 4 plots both total TFP loss and net TFP loss as a function of α. It is clear that the for all values of α, the value of total TFP loss is above its counterpart for net TFP loss. More importantly, the optimal α that minimizes the total TFP loss is 1.60, which is smaller than its counterpart that minimizes the net TFP loss (2.09). This is because credit subsidy, by leading to more inefficient entry and exit, increases fixed capital inputs.

Figure 5 shows that the optimal level of credit subsidy α that minimizes the total TFP loss increases in markup dispersion σ_μ^2. This is because as the markup dispersion σ_μ^2 increases, the magnitude of TFP gain from increasing α preferential credit policy increases, while TFP loss due to the entry and exit effect does not depends on α. Therefore, it is optimal to provide stronger credit support for sectors with higher markups.[16]

5 Conclusion

This paper provides an analysis of the economic rationale behind the preferential credit policies in the presence of market imperfections. Using China's firm-level data, we find empirically that credit subsidy and markup are positively correlated at the sector level, which suggests that the government tends to provide credit subsidy to sectors with high profit margins. We develop a model that incorporates markup heterogeneity and endogenous firm entry and exit to evalute the role of preferential credit policy for aggregate productive efficiency. Our key finding is that sector-specific preferential credit policies trade off the efficiency gain of reducing markup dispersion against the efficiency loss by allowing less productive firms to enter and survive without exiting. Quantitative analysis suggests a positive correlation between the credit subsidy and markup at the sector level as a second-best policy.

[16]Note that when markup dispersion drops to 0.06, optimal α reaches to to its minimum level, 1, the lowest level, which implies no preferential credit subsidy. This is because endogenous firm entry and exit in our model amplifies the efficiency loss due to preferential credit subsidy compared with a static model without endogenous firm entry and exit. Hence, the optimal α would reach to its minimum value, which is one, even before the markup dispersion approaches zero.

Our framework can be extended in several directions. First, the model can incorporate endogenous technological innovation, such that preferential credit subsidy improves not only allocative efficiency, but also efficient level of TFP in the subsidized sectors. Second, the model can endogenize markups by allowing for Kimball demand function at the sectoral level. Since the sectoral average markup increases in credit subsidy, this would potentially enhance optimal credit subsidy towards sectors with higher profit margins.[17] Third, the model can endogenize the demand side of the economy, say by introducing the household sector, and be used to study other types of industrial policies, such as consumer credits, or tax rebates, for specific sectors. We leave these extensions for furture research.

[17] We confirm this result under the two-sector economy with demand for intermediate goods produced by the sector with higher markups subject to Kimball preference. A generalization of our finding to the economy with multiple sectors will be important future work.

Table 1: Summary statistics (4-digit sectoral level)

	Obs.	**Mean**	**Std. Dev.**	**P5**	**P25**	**P50**	**P75**	**P95**
Markup	4188	1.216	0.111	1.050	1.155	1.213	1.272	1.381
Effective interest rate (%)	4188	2.187	0.516	1.467	1.874	2.119	2.443	3.221
Subsidy rate	4188	0.627	0.066	0.530	0.586	0.625	0.664	0.737
Zombie rate	3768	0.359	0.172	0.135	0.227	0.333	0.471	0.667
SOE share (%)	4187	17.319	17.835	0.265	4.017	11.500	25.027	53.658
Export share (%)	4187	16.387	17.236	0.495	3.946	9.568	23.009	53.291

Note: The table presents 4-digit sector-year averages of firm-level data. The "%" symbol denotes percentage terms. "Markup" denotes the average markup of firms. "Effective interest rate (%)" represents the calculated effective interest rate faced by firms (sector-year averages). "Subsidy rate" refers to τ, the calculated subsidy rate as defined in the paper. "Zombie ratio (%)" indicates the percentage of zombie firms at the 4-digit sector-year level. "SOE share (%) " and "Export share (%)" respectively represent the proportion of state-owned capital in total capital and the portion of exports in output (sector-year averages). "Std. Dev." stands for standard deviation, and P5, P25, P50, P75, and P95 represent the 5th, 25th, 50th, 75th, and 95th percentiles, respectively.

Table 2: Estimations of the elasticity of $1 - \tau$ to sectoral markup

	OLS	**Reduced Form**	**First stage**	**Second stage**
Outcome	$y_{i,j,t}$ (1)	$y_{i,j,t}$ (2)	$\bar{\mu}_{i,t}$ (3)	$y_{i,j,t}$ (4)
$\mu_{i,t}$	0.016			-0.648**
	(0.025)			(0.277)
$z_{i,t}^{BI}$		-0.408**	0.631***	
		(0.174)	(0.011)	
Sectoral controls: (SOE share, export share)	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	908,865	908,865	908,865	908,865
Clusters	223,765	223,765	223,765	223,765
R^2	0.018	0.018	0.550	0.018

Note: This table reports the regression estimates of β from the OLS and the IV approach. Standard errors are clustered at the firm level and presented in the parenthesis. *, **, *** denote statistical significance at the 10, 5, and 1 percent levels, respectively.

Table 3: Sectoral zombie ratio and subsidy rate (IV estimations)

	(1)	(2)	(3)	(4)
	Calculated τ		Predicted τ	
	Zombie ratio	Zombie ratio	Zombie ratio	Zombie ratio
Subsidy rate (τ)	2.303***	6.254	1.885***	2.588***
	(0.717)	(6.006)	(0.488)	(0.703)
Sectoral controls: (SOE share, export share)	Yes	Yes	Yes	Yes
Sector FE	Yes	Yes	Yes	Yes
N	3767	3767	3767	3767
R^2	0.807	0.801	0.800	0.800
$adj.R^2$	0.783	0.776	0.775	0.774

Note: This table displays four IV regression estimates on the relationship between the zombie ratio and subsidy rate, using official lending rates as the instrumental variable to address endogeneity. The firms' subsidy rate τ is calculated directly from ASIF data in columns 1 and 2, while in columns 3 and 4, it is predicted τ from Table 2. The sectoral τ are computed as simple averages of firm subsidy rate in columns 1 and 3, and are computed as weighted averages (by output) in columns 2 and 4. The outcome variable is the sectoral zombie ratio. FE = fixed effects. Standard errors are clustered at 4-digit sectoral level. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 4: Parameter Values

Parameters	Interpretation	Source	Value
β	Discount factor for firms	A standard value in the literature	0.96
R	Interest rate	Normalized to 1.04	1.04
η	The elasticity of substitution across sectors	Borrowed from the literature Epifani and Gancia (2011)	5
γ	The tail parameter of Pareto distribution	Match the sales share of top 5% firm, which is 0.4353 in 2006	7.68
ρ	$1-\rho$ is the size of productivity drop	Target the sales share of firms who exited the market in 2006, which is 0.016	0.953
ϕ	$1-\phi$ is the probability of exogenous exit	Target the mass share of new entrants in 2006, which is 0.02	0.986
q	$1-q$ is the probability of entering the low productivity regime	Target the mass share of zombie firms in 2006, which is 0.23	0.761

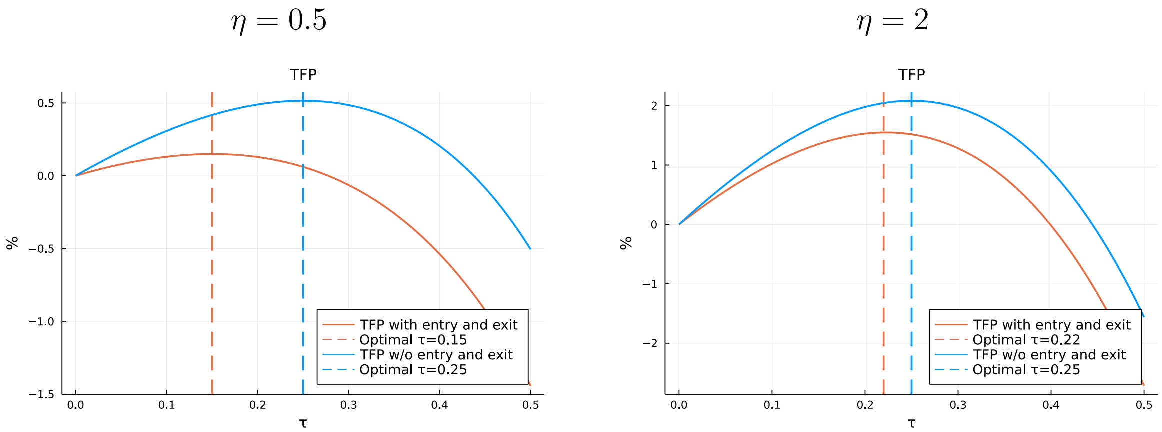


Figure 1: Optimal credit subsidy in a two-sector economy

Note: In each panel, the blue solid line represents the level of log(TFP) in the baseline economy. The red solid line represents the level of log(TFP) in the counterfactual economy without firm entry and exit. For each economy, the level of log(TFP) is normalized by dividing by its value at $\tau = 0$.

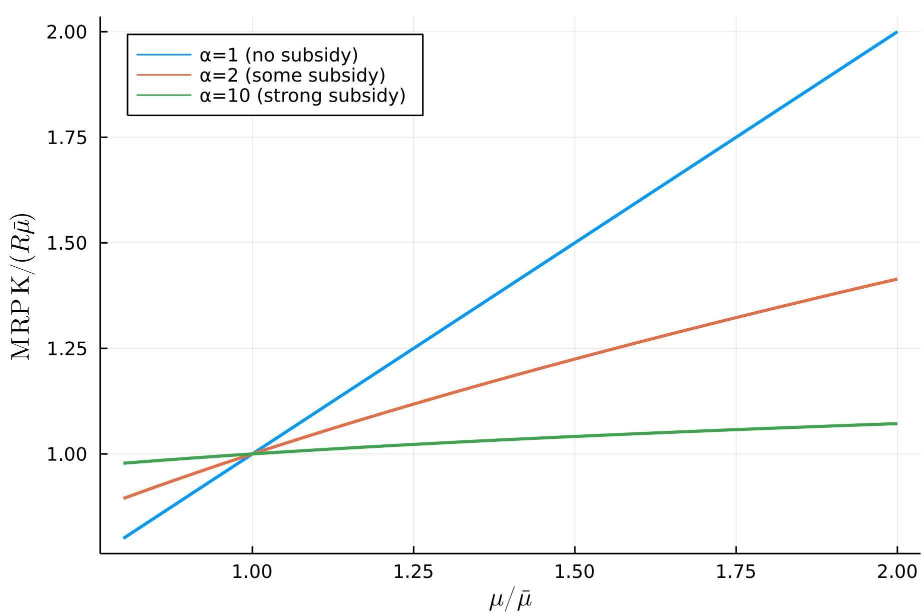


Figure 2: $\mu(1-\tau)/\overline{\mu}$ as a function of $\mu/\overline{\mu}$ under different subsidy policy parameter α

Figure 3: net TFP loss as a function of subsidy policy parameter α when $\eta = 5$

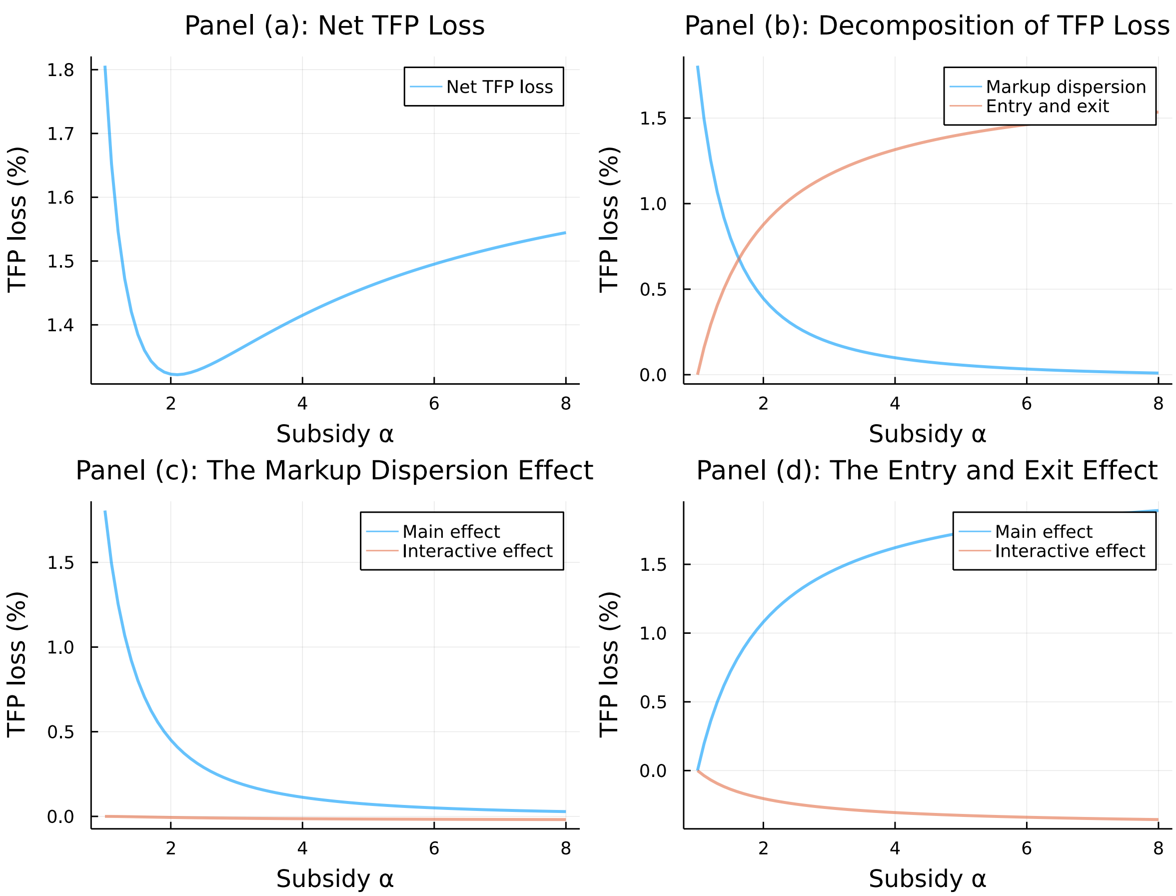
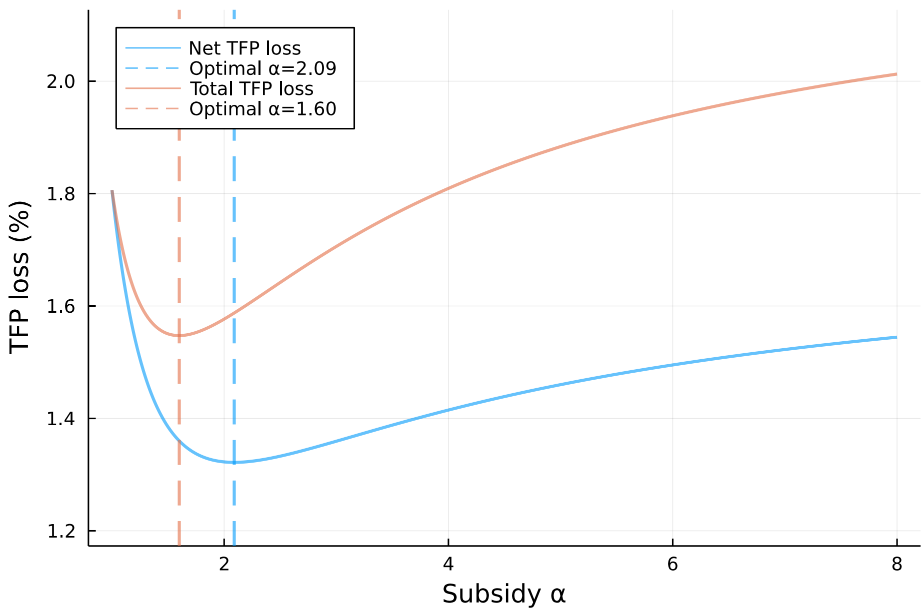
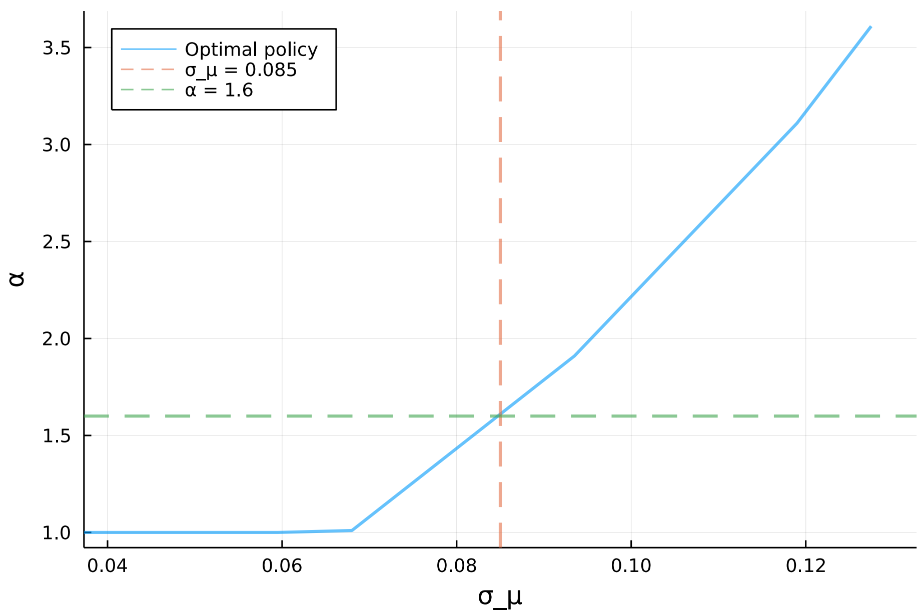


Figure 4: TFP loss as a function of subsidy policy parameter α when $\eta = 5$



Note: The solid red (blue) line plots the total (net) TFP loss when the fixed costs is (not) included in the measured aggregate capital stock. The dash red (blue) lines corresponds to the level of optimal α under these two alternative measures of aggregate TFP.

Figure 5: Optimal α as a function of sectoral markup dispersion



Note: The solid blue line plots the optimal α as a function of $\sigma_m u$. The green (red) dash line corresponds to the level of optimal α in our benchark calibration (the level of standard deviation of sectoral markup).

References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Indentification Properties of Recent Production Estimator." *Econometrica*, 83: 2411—-2451.

Caballero, Ricardo J., Takeo Hoshi, and Anil K. Kashyap. 2008. "Zombie Lending and Depressed Restructuring in Japan." *American Economic Review*, 98: 1943—-1977.

David, Joel, and Venky Venkateswaran. 2019. "The Sources of Capital Misallocation." *American Economic Review*, 109: 2531–2567.

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and Firm-Level Export Status." *American Economic Review*, 102: 2437—-2461.

Dipippo, Gerald, Ilaria Mazzocco, and Scott Kennedy. 2022. "Red Ink: Estimating Chinese Industrial Policy Spending in Competitive Perspective." Center for Strategic and International Studies.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2015. "Competition, markups, and the gains from international trade." *American Economic Review*, 105(10): 3183–3221.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2023. "How costly are markups?" *Journal of Political Economy*, 131(7): 1619–1675.

Epifani, Paolo, and Gino Gancia. 2011. "Trade, markup heterogeneity and misallocations." *Journal of International Economics*, 83: 1–13.

Fukuda, Shinichi, and Junichi Nakamura. 2011. "Why Did 'Zombie' Firms Recover in Japan?" *The World Economy*, 37: 1124–1137.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik Instruments: What, When, Why, and How." *American Economic Review*, 110: 2586—-2624.

Guner, Nezih, Gustavo Ventura, and Yi Xu. 2008. "Macroeconomic Implications of Size Dependent Policies." *Review of Economic Dynamics*, 11: 721—-744.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, CXXIV(4): 1403–1448.

Huang, S., and Y. Chen. 2017. "The distribution features and classified disposition of China's zombie firms." *China Industrial Economics*, 03: 24–43.

Kim, Minho, Munseob Lee, and Yongseok Shin. 2021. "The Plant-Level View of an Industrial Policy: The Korean Heavy Industry Drive of 1973." Washington University in St. Louis.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies*, 70: 317—-341.

Lin, Yifun, Mingxing Liu, and Qi Zhang. 2004. "Policy burden and enterprise's soft budgetary binding: a case study from China." *Management World*, 8(1): 81–89.

Liu, Ernest. 2019. "Industrial Policies in Production Network." *Quarterly Journal of Economics*, 134: 1883–1948.

Lu, Yi, and Linhui Yu. 2015. "Trade Liberalization and Markup Dispersion: Evidence from China's WTO Accession." *American Economic Journal: Applied Economics*, 7: 221–253.

Midrigan, Virgiliu, and Daniel Yi Xu. 2014. "Finance and Misallocation: Evidence from Plant-Level Data." *American Economic Review*, 104(2): 422–58.

Olley, Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunication Equipment." *Econometrica*, 64: 1264—-1273.

Restuccia, Diego, and Richard Rogerson. 2008. "Policy distortions and aggregate productivity with heterogeneous establishments." *Review of Economic dynamics*, 11(4): 707–720.

Restuccia, Diego, and Richard Rogerson. 2017. "The Causes and Costs of Misallocation." *Journal of Economic Perspectives*, 31: 151—-174.

Tan, Yuyan, Zhibo Tan, Yiping Huang, and Wing Tye Woo. 2017. "The Crowding-out Effect of Zombie Firms: Evidence from China's Industrial Firms." *Economic Research Journal (in Chinese)*, 5: 322–364.

Wu, Guiying Laura. 2018. "Capital Misallocation in China: Financial Frictions or Policy Distortions?" *Journal of Development Economics*, 130: 203—-223.

Yang, Mu-Jeung. 2021. "Micro-Level Misallocation and Selection." *American Economic Journal: Macroeconomics*, 13: 341–368.

Internet Appendices

Not Intended for Publication

A Estimation of firm-level TFP

To obtain the 4-digit sectoral TFPs, we first estimate the firm-level TFP following Levinsohn and Petrin (2003) by assuming the following Cobb-Douglas production technology

$$y_t = \beta_0 + \beta_l l_t + \beta_k k_t + \omega_t + \eta_t \tag{S1}$$

where all lowercase letters denote the loge y_t is the logarithms of the variables. y_t is the firm's output, measured as the value added. labor l_t is the free variable input, and k_t is the state variable capital. The error term is assumed to be two additive terms: a transmitted productivity component ω_t and an i.i.d. component η. Only the former is an unobserved state variable impacting the firm's decision. Ignoring the endogeneity problem will lead to inconsistent results in the production function estimation.

We use the firm's intermediate input m_t as the proxy variable, which depends on the state variables k_t and ω_t:

$$m_t = m_t(k_t, \omega_t) \tag{S2}$$

Following the assumption in Levinsohn and Petrin (2003), the demand function is monotonically increasing in ω_t. This implies the inversion of the input demand function

$$\omega_t = \omega_t(k_t, m_t) \tag{S3}$$

We assume the unobserved productivity w_t follows a first-order Markov process and capital k_t does not immediately respond to ϵ_t (Olley and Pakes, 1996; Levinsohn and Petrin, 2003).

$$\omega_t = E[\omega_t|\omega_{t-1}] + \epsilon_t \tag{S4}$$

Given the input demand function, the production function can be written as

$$y_t = \beta_0 + \beta_l l_t + \beta_k k_t + \omega_t + \eta_t \tag{S5}$$

$$= \beta_l l_t + \phi_t(k_t, m_t) + \eta_t \tag{S6}$$

where

$$\phi_t(k_t, m_t) = \beta_0 + \beta_k k_t + \omega_t(k_t, m_t) \tag{S7}$$

To complete the first stage estimation in Levinsohn and Petrin (2003), we substituting a third-order polynomial approximation in k_t and m_t in place of $\phi_t(k_t, m_t)$ to consistently estimate the value-added equation (using OLS):

$$y_t = \delta_0 + \beta_l l_t + \sum_{i=0}^{3}\sum_{j=0}^{3-i} \delta_{ij} k_t^i m_t^j + \eta_t \tag{S8}$$

where we can obtain the estimates of β_l and ϕ_t. We identify the coefficient β_k from the second stage of the estimation routine, which begins by first computing the estimated value for ϕ_t:

$$\widehat{\phi_t} = \widehat{y_t} - \widehat{\beta_l} l_t \tag{S9}$$

$$= \widehat{\delta_0} + \sum_{i=0}^{3}\sum_{j=0}^{3-i} \widehat{\delta_{ij}} k_t^i m_t^j - \widehat{\beta_l} l_t \tag{S10}$$

For any given value β_k^*, we can compute a prediction for ω_t for all time t using

$$\widehat{\omega_t} = \widehat{\phi_t} - \beta_k^* k_t \tag{S11}$$

$E[\omega_t|\omega_{t-1}]$ is then given by the predicted value from the regression:

$$E[\widehat{\omega_t|\omega_{t-1}}] = \gamma_0 + \gamma_1 \omega_{t-1} + \gamma_2 \omega_{t-1}^2 + \gamma_3 \omega_{t-1}^3 + \epsilon_t \tag{S12}$$

The second stage has only one parameter to be estimated, β_k, which is defined as the solution to

$$\min_{\beta_k^*} \sum_t (y_t - \widehat{\beta_l} l_t - \beta_k^* k_t - E[\widehat{\omega_t|\omega_{t-1}}])^2 \tag{S13}$$

A bootstrap approach is used to construct standard errors for $\widehat{\beta_l}$ and $\widehat{\beta_k}$. The firm-level TFP is presented in Eq.(14).

$$TFP_{i,t} = \exp\left(y_{i,t} - \widehat{\beta_l} l_{i,t} - \widehat{\beta_k} k_{i,t}\right) \tag{S14}$$

The 4-digit sectoral TFP is calculated as the weighted average of the firm-level TFP.

$$TFP_{s,t} = \frac{\sum_i (y_{i,t} TFP_{i,t})}{\sum_i y_{i,t}} \tag{S15}$$

B Estimation of markups

B.1 The firm-level markup

In our model, capital is the only factor of production. But in comparing our model's implications for markups to the data, it is important to recognize that, factor shares such as labor and intermediate inputs also differ across producers. To control for such sources of heterogeneity, we relax the assumptions of a single factor of production when mapping our model into firm-level production data. In the following, we illustrate our empirical estimation of firm-level markups, and how we aggregate them to obtain sector-level markups.

The production function of firm i at time t is given by:

$$Y_{i,t} = F_{i,t}(X_{i,t}, a_{i,t}), \tag{S16}$$

where $Y_{i,t}$ is firm's real output level, $X_{i,t}$ represents the physical inputs, including capital $K_{i,t}$, labor $L_{i,t}$, and intermediate inputs $M_{i,t}$, $a_{i,t}$ denotes firm-specific productivity. The production function $F_{i,t}$ is assumed to be continuous and twice-differentiable with respect to all inputs.

The cost-minimization problem faced by firm i at time t is:

$$\min_{L_{i,t},K_{i,t},M_{i,t}} w_{i,t}L_{i,t} + r_{i,t}K_{i,t} + p^m_{i,t}M_{i,t} \tag{S17}$$

$$s.t. \quad F_{i,t}(L_{i,t}, K_{i,t}, M_{i,t}, a_{i,t}) \geq \bar{Y}_{i,t}, \tag{S18}$$

where $w_{i,t}$, $K_{i,t}$, and $M_{i,t}$ are the wage rate, capital rental rate and prices of intermediate inputs. $\bar{Y}_{i,t}$ denotes the minimum level of output. The Lagrangian function associated with cost-minimization problem can be written as:

$$\mathcal{L}(L_{i,t}, K_{i,t}, M_{i,t}, \lambda_{i,t}) = w_{i,t}L_{i,t} + r_{i,t}K_{i,t} + p^m_{i,t}M_{i,t} + \lambda_{i,t}[\bar{Y}_{i,t} - F_{i,t}(L_{i,t}, K_{i,t}, M_{i,t}, a_{i,t})] \tag{S19}$$

In general, any factor inputs $X_{i,t}$ can be used for markup calculation. Since capital is often considered to be a dynamic input, and labor is often not freely chosen in China, especially for the state-owned enterprises (SOEs) due to their implicit policy burdens (Lu and Yu, 2015), we calculate the markup based on firms' optimal choice of intermediate inputs. The first-order conditions for intermediate inputs:

$$\frac{\partial \mathcal{L}}{\partial M_{i,t}} = p^m_{i,t} - \lambda_{i,t}\frac{\partial F_{i,t}}{\partial M_{i,t}} = 0, \tag{S20}$$

where $\lambda_{i,t}$ represent the marginal cost of production. Rearranging equation (S20) and multiplying both sides by $\frac{M_{i,t}}{Y_{i,t}}$:

$$\frac{\partial F_{i,t}}{\partial M_{i,t}}\frac{M_{i,t}}{Y_{i,t}} = \frac{1}{\lambda_{i,t}}\frac{p^m_{i,t}M_{i,t}}{Y_{i,t}} = \frac{P_{i,t}}{\lambda}\frac{p^m_{i,t}M_{i,t}}{P_{i,t}Y_{i,t}}, \tag{S21}$$

where $P_{i,t}$ is price of final goods. The firm-level markup can be defined as:

$$\mu_{i,t} = \frac{P_{i,t}}{\lambda_{i,t}} = \theta^m_{i,t}\left(\frac{p^m_{i,t}M_{i,t}}{P_{i,t}Y_{i,t}}\right)^{-1}, \tag{S22}$$

where, $\theta^m_{i,t} \equiv \frac{\partial F_{i,t}}{\partial M_{i,t}}\frac{M_{i,t}}{Y_{i,t}}$ is the output elasticity of intermediate inputs, $\left(\frac{p^m_{i,t}M_{i,t}}{P_{i,t}Y_{i,t}}\right)^{-1}$ is the intermediate inputs cost as a share of total revenue. As shown in equation (S22), the firm-level markup can be calculated using the factor output elasticity and its cost share.

B.2 Markup estimation

The cost share of intermediate inputs $\left(\frac{p^m_{i,t}M_{i,t}}{P_{i,t}Y_{i,t}}\right)$ could be directly calculated using the information in the dataset (expenditure on intermediate materials and sales revenue). To obtain the output elasticity of intermediate input $\theta^m_{i,t}$, we estimate the production function at the 4-digit sector level. Following De Loecker and Warzynski (2012), we assume a translog production function

$$\begin{aligned} y_{i,t} = a_{i,t} &+ \beta_l l_{i,t} + \beta_k k_{i,t} + \beta_m m_{i,t} + \beta_{ll} l^2_{i,t} + \beta_{kk} k^2_{i,t} + \beta_{mm} m^2_{i,t} \\ &+ \beta_{lk} l_{i,t} k_{i,t} + \beta_{km} k_{i,t} m_{i,t} + \beta_{lm} l_{i,t} m_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{S23}$$

where $y_{i,t}$, $l_{i,t}$, $k_{i,t}$, and $m_{i,t}$ represent the logarithm of firms' output, labor (numbers of employees), capital stock, and intermediate inputs. $a_{i,t}$ and $\epsilon_{i,t}$ are the firm's logged productivity and the error term.

Using firm-level data from ASIF, we estimate the translog production function (equation 8) at the two-digit level for each sector separately. After obtaining the estimates of $\{\hat{\beta}_l, \hat{\beta}_k, \hat{\beta}_m, \hat{\beta}_{ll}, \hat{\beta}_{kk}, \hat{\beta}_{mm}, \hat{\beta}_{lk}, \hat{\beta}_{km}, \hat{\beta}_{lm}\}$, The output elasticity of intermediate materials can be calculated as:

$$\hat{\theta}_t^m = \frac{\partial y_{i,t}}{\partial m_{i,t}} = \hat{\beta}_m + 2\hat{\beta}_{mm} m_{i,t} + \hat{\beta}_{lm} l_{i,t} + \hat{\beta}_{km} k_{i,t} \tag{S24}$$

In order to obtain consistent estimates of the output elasticity, we mainly follow the procedure in De Loecker and Warzynski (2012), which relies on proxy methods provided by Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg, Caves and Frazer (2015) to deal with the simultaneity problem caused by the correlations between productivity and input choices.

Following the timing assumptions and rationale in Ackerberg, Caves and Frazer (2015), a firm's intermediate input demand can be written as a function of its capital, labor, and productivity:

$$m_{i,t} = f(k_{i,t}, l_{i,t}, a_{i,t}) \tag{S25}$$

$$a_{i,t} = g(k_{i,t}, l_{i,t}, m_{i,t}) \tag{S26}$$

Combining with equation (S8), the production function can be re-written as:

$$\begin{aligned} y_{i,t} =& g(k_{i,t}, l_{i,t}, m_{i,t}) + \beta_l l_{i,t} + \beta_k k_{i,t} + \beta_m m_{i,t} + \beta_{ll} l_{i,t}^2 + \beta_{kk} k_{i,t}^2 + \beta_{mm} m_{i,t}^2 \\ &+ \beta_{lk} l_{i,t} k_{i,t} + \beta_{km} k_{i,t} m_{i,t} + \beta_{lm} l_{i,t} m_{i,t} + \epsilon_{i,t} \end{aligned} \tag{S27}$$

or simply

$$y_{i,t} = h(k_{i,t}, l_{i,t}, m_{i,t}) + \epsilon_{i,t}, \tag{S28}$$

In the first stage of estimation, we can estimate equation (S27) and obtain the estimates of $h(k_{i,t}, l_{i,t}, m_{i,t})$ and $\epsilon_{i,t}$. The logged productivity $a_{i,t}$ can then be obtained given some candidate parameter vector $\boldsymbol{\beta} = (\hat{\beta}_l, \hat{\beta}_k, \hat{\beta}_m, \hat{\beta}_{ll}, \hat{\beta}_{kk}, \hat{\beta}_{mm}, \hat{\beta}_{lk}, \hat{\beta}_{km}, \hat{\beta}_{lm})$[18].

$$\begin{aligned} a_{i,t}(\boldsymbol{\beta}) = h_{i,t}(\cdot) &- \beta_l l_{i,t} - \beta_k k_{i,t} - \beta_m m_{i,t} - \beta_{ll} l_{i,t}^2 - \beta_{kk} k_{i,t}^2 - \beta_{mm} m_{i,t}^2 \\ &- \beta_{lk} l_{i,t} k_{i,t} + \beta_{km} k_{i,t} m_{i,t} - \beta_{lm} l_{i,t} m_{i,t} \end{aligned} \tag{S29}$$

In the second stage, we estimate all production function coefficients following the standard assumption in the literature (De Loecker and Warzynski, 2012; Edmond, Midrigan and Xu, 2015). The logged productivity follows a flexible AR(1) process in equation (S30), where $\phi(\cdot)$ is a second-order polynomial.

[18]For example, we use the OLS estimates of equation (S27)

$$a_{i,t}(\boldsymbol{\beta}) = \phi(a_{i,t-1}(\boldsymbol{\beta})) + \xi_{i,t}(\boldsymbol{\beta}) \tag{S30}$$

Through the non-parametric regression of $a_{i,t}(\boldsymbol{\beta})$ on its lag $a_{i,t-1}(\boldsymbol{\beta})$, we can recover the idiosyncratic shock to productivity given $\boldsymbol{\beta}$. To obtain the estimates of production function parameter $\boldsymbol{\beta}$, we apply standard GMM estimation relying on the sequential exogeneity that $\boldsymbol{\beta}$ is uncorrelated with the lag of inputs. Following Ackerberg, Caves and Frazer (2015), the capital stock is assumed to be a pre-determined variable and therefore enters without a lag. The moment condition is then given by equation (S31). After obtaining the estimates of $\boldsymbol{\beta}$, we can calculate the output elasticity using equation (S24) and subsequently the firm markup using equation (S22).

$$E\left(\xi_{i,t}(\boldsymbol{\beta}) z_{i,t}\right) = 0, \tag{S31}$$

where $z_{i,t}$ is given by:

$$z_{i,t} = \left[l_{i,t-1}, k_{i,t}, m_{i,t-1}, l^2_{i,t-1}, k^2_{i,t}, m^2_{i,t-1}, l_{i,t-1}k_{i,t}, l_{i,t-1}m_{i,t-1}, k_{i,t}m_{i,t-1}\right] \tag{S32}$$

C Identification of Zombie Firms in Firm-level Data

Zombie firms are identified based on their ability to meet their minimum interest rate payments. This involves constructing a hypothetical minimum interest payment and comparing it with the actual interest payment or earnings of the firm (Caballero, Hoshi and Kashyap, 2008; Fukuda and Nakamura, 2011). We primarily follow this approach in identifying zombie firms and use both methods outlined in Tan et al. (2017) and Huang and Chen (2017) when calculating the hypothetical minimum interest rate payments. In addition, we follow recent studies that identify zombie firms in China using list firm data. We remove observations from the zombie firms' group if (i) they are only identified as zombies in one year but not the year before and the year after. (ii) they have had a growing net asset for three consecutive years. The detailed steps are as follows:

1. Calculate the minimum interest payment $R^*_{i,t}$ that enterprise i should be required to pay in year t under normal operation:

$$R^*_{i,t} = rs_{t-1}BS_{i,t-1} + \left(\frac{1}{5}\sum_{j=1}^{5} rl_{t-j}\right) BL_{(i,t-1)}$$

   where $BS_{i,t-1}$ and $BL_{(i,t-1)}$ are short-term and long-term bank loans, respectively. As there is no directly detailed bank liability in the database, we use short-term borrowing, which is the difference between the enterprise's short-term liabilities and accounts and other payable as the firm short-term borrowing [19], and uses the long-term liabilities as long-term bank borrowing. rs_{t-1} and rl_{t-j} are 0.9 times the annual one-year and five-year average benchmark loan interest rates of a bank in year t, respectively[20].

2. Estimate the firm's interest income. Since only the actual net interest expense of the firms (calculated as interest expense minus interest income), $RN_{i,t}$ is available in the database, in order to compare the actual interest expenditure of the firms with the benchmark interest expense, it is necessary to first estimate the enterprise's interest income $RI_{i,t}$:

$$RI_{i,t} = (CA_{i,t-1} - AR_{i,t-1} - INV_{i,t-1}) * rd_t$$

where $CA_{i,t-1}$, $AR_{i,t-1}$, and $INV_{i,t-1}$ represent the enterprise's current assets, accounts receivable, and inventory, respectively, and rd_t represents the one-year benchmark deposit interest rate in year t.

3. Compare the enterprise's actual net interest expenditure $RN_{i,t}$ with the calculated minimum net interest expenditure $(R^*_{i,t} - RI_{i,t})$, and standardize the difference using the previous period's borrowing $B_{i,t-1} = BS_{i,t-1} + BL_{i,t-1}$ to obtain the interest gap, $gap_{i,t}$:

$$gap_{i,t} = \left(RN_{i,t} - (R^*_{i,t} - RI_{i,t})\right) / B_{i,t-1}$$

According to Caballero, Hoshi and Kashyap (2008), if $gap_{i,t} < 0$, it indicates that the firm has received subsidies and is classified as a zombie firm; otherwise, the firm is not a zombie firm. Furthermore, Fukuda and Nakamura (2011) believe that the above measurement method is more likely to classify enterprises with good operating performance and low financing costs as zombie firms. Therefore, further correction is made using the information on the firm's profit:

$$gapadj_{i,t} = \left(EBIT_{i,t} - \left(R^*_{i,t} - RI_{i,t}\right)\right) / B_{i,t-1}$$

[19]Based on data availability, the operating liabilities include accounts payable, value-added tax payable, income tax payable, wages payable, and employee benefits payable. Missing data were adjusted based on payable items from other years of the company. We consider the year-end balance of payable income tax to be 1/4 of the cumulative amount (quarterly payment) and the year-end balance of value-added tax payable, wages payable, and employee benefits payable to be 1/12 of the cumulative amount (monthly payment)

[20]During the period from 1998 to 2011, the People's Bank of China stipulated that the lower limit of the floating range of loan interest rates for financial institutions should be 0.9 times the benchmark interest rate.

If the firm's earnings before interest and taxes are greater than the minimum net interest expense ($gapadj_{i,t} > 0$), the firm is changed to a non-zombie firm.

Although many financial characteristics of the firms are considered, the above classification method may not be accurate enough. As stated by Lin, Liu and Zhang (2004), in China, if a company is unable to repay its loans, the bank will often roll over the overdue payments and interest into a new bank loan for the following year. Therefore, we follow the suggestions in Fukuda and Nakamura (2011) and Tan et al. (2017), and further modify the measurement. A firm is identified as a zombie firm in period t if it satisfies all the following three conditions: (i) the pre-tax profit of the firm is less than the minimum net interest expense. (ii) the firm's liabilities exceed 50% of its total assets, (iii) the liabilities in period t are greater than those in period $t-1$.

After identifying whether a firm is zombie or not, we can calculate the ratio of zombie firms at the 4-digit sectoral level by dividing the total number of zombie firms by the total number of firms in the sector. We also attempt to calculate the sectoral zombie ratio as the sum of the fixed assets of zombie firms divided by the sum of fixed assets of all firms in the sector. Our regression results remain robust regardless of which method we use to calculate the sectoral zombie ratio.

D Proof of Propositions and Lemmas

Derivation of Equation (3.3)

$$\begin{aligned}
\Pi_s &= \int_{i\in I_s} \pi_s(a_s(i), \tau_s)di \\
&= N_s \int_{a=\underline{a}_s}^{\infty} \pi_s(a, \tau_s)d\tilde{G}(a) \\
&= N_s \int_{a=\underline{a}_s}^{\infty} \left(\Lambda_s \cdot \left(\frac{a}{R(1-\tau_s)}\right)^{\varepsilon_s - 1} - c_{fs}\right) d\tilde{G}(a) \\
&= N_s \left(\Lambda_s \cdot \left(\frac{\left(\int_{a=\underline{a}_s}^{\infty} a^{\varepsilon_s - 1} d\tilde{G}(a)\right)^{\frac{1}{\varepsilon_s - 1}}}{R(1-\tau_s)}\right)^{\varepsilon_s - 1} - c_{fs}\right) \\
&= N_s \left(\Lambda_s \cdot \left(\frac{\bar{a}_s}{R(1-\tau_s)}\right)^{\varepsilon_s - 1} - c_{fs}\right) \\
&= N_s \pi_s(\bar{a}_s, \tau_s)
\end{aligned}$$

Proof of Lemma 1. Using Equation (11) and (9), we have $K_s = \int_{i\in I_s} k_s(i)\,di = \int_{i\in I_s} \frac{y_s(i)}{a_s(i)}\,di = \int_{i\in I_s} \frac{y_s}{N_s a_s(i)}\left(\frac{p_s(i)}{p_s}\right)^{-\varepsilon_s} di$. According to Equation (12) and (19), we have $\frac{p_s(i)}{p_s} = \frac{\bar{a}_s}{a_s(i)}$. Thus, $K_s = \frac{y_s}{N_s}(\bar{a}_s)^{-\varepsilon_s}\int_{i\in I_s} a_s(i)^{\varepsilon_s-1}\,di = \frac{y_s}{N_s}(\bar{a}_s)^{-\varepsilon_s} N_s \bar{a}_s^{\varepsilon_s-1}\,di = \frac{y_s}{\bar{a}_s}$. ∎

Proof of Proposition 1. Taking log with respect to (30), we have

$$\log(\text{TFP}^n) = \log\left(\sum_{s=1}^{S}(\mu_s)^{\frac{1-\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}(a_s)^{\eta-1}\right)^{\frac{\eta}{\eta-1}} - \log\sum_{s=1}^{S}(\mu_s)^{-\frac{\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}(a_s)^{\eta-1} \quad \text{(S33)}$$

We solve the first and second arguments on the right-hand-side of Equation ($S33$) separately. For the first argument, we have

$$\begin{aligned}
&\log\left(\sum_{s=1}^{S}(\mu_s)^{\frac{1-\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}(a_s)^{\eta-1}\right)^{\frac{\eta}{\eta-1}}\\
= \;& \frac{\eta}{\eta-1}\log\mathbb{E}\exp\left(\log(\mu_s)^{\frac{1-\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}} + \log(a_s)^{\eta-1}\right)\\
= \;& \frac{\eta}{\eta-1}\left[\begin{array}{c}\mathbb{E}\log(\mu_s)^{\frac{1-\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}} + \mathbb{E}\log(a_s)^{\eta-1} + Cov\left(\log(\mu_s)^{\frac{1-\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}, \log(a_s)^{\eta-1}\right)\\ +\frac{1}{2}Var\left(\log(\mu_s)^{\frac{1-\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}\right) + \frac{1}{2}Var\left(\log(a_s)^{\eta-1}\right)\end{array}\right]\\
= \;& \frac{\eta}{\eta-1}\left[\log\mathbb{E}(\mu_s)^{\frac{1-\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}} + \log\mathbb{E}(a_s)^{\eta-1} + Cov\left(\log(\mu_s)^{\frac{1-\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}, \log(a_s)^{\eta-1}\right)\right] \quad \text{(S34)}
\end{aligned}$$

The second equality uses the property that

$$\begin{aligned}
\log(z) &= \log\sum f_1(x)\, f_2(y)\,\Phi(x,y)\\
&= \log\mathbb{E}\left[\exp\left(\log f_1(x) + \log f_2(y)\right)\right]\\
&= \mathbb{E}\log f_1(x) + \mathbb{E}\log f_2(y) + \frac{1}{2}Var\left[f_1(x)\right] + \frac{1}{2}Var\left[f_2(y)\right] + Cov\left(f_1(x), f_2(y)\right)
\end{aligned}$$

The third equality follows that

$$\mathbb{E}\log(\mu_s)^{\frac{1-\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}} = \log\mathbb{E}\left((\mu_s)^{\frac{1-\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}\right) - \frac{1}{2}Var\left(\log(\mu_s)^{\frac{1-\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}\right)$$

Similarly, for the second argument, we have

$$\begin{aligned}
& \log\sum_{s=1}^{S}(\mu_s)^{-\frac{\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}(a_s)^{\eta-1} \\
= \; & \log\mathbb{E}\exp\left(\log(\mu_s)^{-\frac{\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}+\log(a_s)^{\eta-1}\right) \\
= \; & \mathbb{E}\log(\mu_s)^{-\frac{\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}+\mathbb{E}\log(a_s)^{\eta-1}+Cov\left(\log(\mu_s)^{-\frac{\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}},\log(a_s)^{\eta-1}\right) \\
& +\frac{1}{2}Var\left(\log(\mu_s)^{-\frac{\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}\right)+\frac{1}{2}Var\left(\log(a_s)^{\eta-1}\right) \\
= \; & \log\mathbb{E}(\mu_s)^{-\frac{\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}}+\log\mathbb{E}(a_s)^{\eta-1}+Cov\left(\log(\mu_s)^{-\frac{\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}},\log(a_s)^{\eta-1}\right) \quad \text{(S35)}
\end{aligned}$$

We then solve for the two covariance terms

$$\frac{\eta}{\eta-1}Cov\left(\log(\mu_s)^{\frac{1-\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}},\log(a_s)^{\eta-1}\right)=\frac{\eta\,(\eta-1)\,(1-\alpha-\gamma)}{\alpha\gamma}Cov\left(\log(\mu_s),\log(a_s)\right)$$

$$Cov\left(\log(\mu_s)^{-\frac{\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}},\log(a_s)^{\eta-1}\right)=\frac{-\eta\gamma+(1-\alpha)\,(\eta-1)}{\alpha\gamma}\,(\eta-1)\,Cov\left(\log(\mu_s),\log(a_s)\right)$$

Hence,

$$\begin{aligned}
& \frac{\eta}{\eta-1}Cov\left(\log(\mu_s)^{\frac{1-\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}},\log(a_s)^{\eta-1}\right)-Cov\left(\log(\mu_s)^{-\frac{\eta}{\alpha}+\frac{(1-\alpha)(\eta-1)}{\alpha\gamma}},\log(a_s)^{\eta-1}\right) \\
= \; & \frac{(\eta-1)\left[\eta\,(1-\alpha-\gamma)+\eta\gamma-(1-\alpha)\,(\eta-1)\right]}{\alpha\gamma}Cov\left(\log(\mu_s),\log(a_s)\right) \\
= \; & -\left(1-\frac{1}{\alpha}\right)\frac{\eta-1}{\gamma}Cov\left(\log(\mu_s),\log(a_s)\right) \qquad \text{(S36)}
\end{aligned}$$

Plugging $(S34)$, $(S35)$ and $(S36)$ back to $(S33)$ and reordering, we have (31). ■

Derivation of Equation (35). Note that Equation (14) implies

$$N_s=(\mu_s-1)\mu_s^{-\varepsilon_s}p_s^{\varepsilon_s}y_s\underline{a}_s^{\varepsilon_s-1}(R(1-\tau_s))^{1-\varepsilon_s}c_{fs}^{-1}.$$

Therefore, we have

$$\begin{aligned}
N_s\bar{k} &= (\mu_s-1)\mu_s^{-\varepsilon_s}p_s^{\varepsilon_s}y_s\underline{a}_s^{\varepsilon_s-1}(R(1-\tau_s))^{1-\varepsilon_s}\frac{1}{R(1-\tau_s)\bar{k}}\bar{k} \quad \left(\text{using } c_{fs}=R(1-\tau_s)\bar{k}\right) \\
&= (\mu_s-1)\mu_s^{-\varepsilon_s}p_s^{\varepsilon_s}\left(\frac{p_s}{P}\right)^{-\eta}Y\underline{a}_s^{\varepsilon_s-1}(R(1-\tau_s))^{-\varepsilon_s} \quad \left(\text{using } y_s=\left(\frac{p_s}{P}\right)^{-\eta}Y\right) \\
&= (\mu_s-1)\mu_s^{-\varepsilon_s}p_s^{\varepsilon_s-\eta}\underline{a}_s^{\varepsilon_s-1}(R(1-\tau_s))^{-\varepsilon_s}P^{\eta}Y.
\end{aligned}$$

Further, we have $K_s = \frac{y_s}{\bar{a}_s} = \left(\frac{p_s}{P}\right)^{-\eta} Y(\bar{a}_s)^{-1}$. Combining them together, we have

$$\begin{aligned}
\frac{\sum_s N_s \bar{k}}{\sum_s K_s} &= \frac{\sum_s (\mu_s - 1)\mu_s^{-\varepsilon_s} p_s^{\varepsilon_s - \eta} \underline{a}_s^{\varepsilon_s - 1}(R(1-\tau_s))^{-\varepsilon_s} P^\eta Y}{\sum_s p_s^{-\eta}(\bar{a}_s)^{-1} P^\eta Y} \\
&= \frac{\sum_s (\mu_s - 1)\mu_s^{-\varepsilon_s} p_s^{\varepsilon_s - \eta} \underline{a}_s^{\varepsilon_s - 1}(R(1-\tau_s))^{-\varepsilon_s}}{\sum_s p_s^{-\eta}(\bar{a}_s)^{-1}} \\
&= \frac{\sum_s (\mu_s - 1)\mu_s^{-\varepsilon_s} \left(\frac{\mu_s R(1-\tau_s)}{\bar{a}_s}\right)^{\varepsilon_s - \eta} \underline{a}_s^{\varepsilon_s - 1}(R(1-\tau_s))^{-\varepsilon_s}}{\sum_s \left(\frac{\mu_s R(1-\tau_s)}{\bar{a}_s}\right)^{-\eta} (\bar{a}_s)^{-1}} \qquad \left(\text{using } p_s = \frac{\mu_s R(1-\tau_s)}{\bar{a}_s}\right) \\
&= \frac{\sum_s (\mu_s - 1)\mu_s^{-\eta}(R(1-\tau_s))^{-\eta} \bar{a}_s^{\eta - \varepsilon_s} \underline{a}_s^{\varepsilon_s - 1}}{\sum_s \mu_s^{-\eta}(R(1-\tau_s))^{-\eta} \bar{a}_s^{\eta - 1}} \\
&= \frac{\sum_s (\mu_s - 1)\mu_s^{-\eta}(R(1-\tau_s))^{-\eta} \left(\frac{\gamma}{\gamma - \varepsilon_s + 1}\right)^{\frac{\eta-1}{\varepsilon_s - 1} - 1} \underline{a}_s^{\eta - 1}}{\sum_s \mu_s^{-\eta}(R(1-\tau_s))^{-\eta} \left(\frac{\gamma}{\gamma - \varepsilon_s + 1}\right)^{\frac{\eta-1}{\varepsilon_s - 1}} \underline{a}_s^{\eta - 1}} \qquad \left(\text{using } \bar{a}_s = \left(\frac{\gamma}{\gamma - \varepsilon_s + 1}\right)^{\frac{1}{\varepsilon_s - 1}} \underline{a}_s\right) \\
&= \frac{\sum_s \mu_s^{-\eta}(R(1-\tau_s))^{-\eta} \left(\frac{\gamma}{\gamma - \varepsilon_s + 1}\right)^{\frac{\eta}{\varepsilon_s - 1} - 1} \underline{a}_s^{\eta - 1} \frac{(\mu_s - 1)(\gamma - \varepsilon_s + 1)}{\gamma}}{\sum_s \mu_s^{-\eta}(R(1-\tau_s))^{-\eta} \left(\frac{\gamma}{\gamma - \varepsilon_s + 1}\right)^{\frac{\eta-1}{\varepsilon_s - 1}} \underline{a}_s^{\eta - 1}} \qquad \text{(S37)}
\end{aligned}$$

Note that the only source of sectoral heterogeneity is the markup μ_s. When all sectors are the same, we have $\frac{\sum_s N_s \bar{k}}{\sum_s K_s} = \frac{(\mu - 1)(\gamma - \varepsilon + 1)}{\gamma}$, which is independent of $\bar{k}$, c_e, and α.

E Calibrating the shape parameter of Pareto distribution of firm-level productivity

This Appendix shows how to use the data about the distribution of revenue shares to calibrate the shape parameter of the Pareto distribution of firm-level productivity

The revenue of firm i in sector s is

$$p_s(i) y_s(i) = \frac{y_s}{N_s} p_s^{-\varepsilon_s} \mu_s \left((1 - \tau_s^K) R\right)^{1 - \varepsilon_s} a_s(i)^{\varepsilon_s - 1} = \Omega_s a_s(i)^{\varepsilon_s - 1}.$$

Let a_s^q be the q-th quantile of the productivity distribution[21]. Then the top (100-q)% (in terms of scale) firms' revenue share in sector s can be represented as

$$Share_s(q) = \frac{\int_{i \in I_s(q)} p_s(i) y_s(i) di}{\int_{i \in I_s} p_s(i) y_s(i) di},$$

[21] that is, $\tilde{G}(a_s^q) = \frac{q}{100}$.

where $I_s(q) \equiv \{i : a_s(i) \geq a_s^q\}$ is the set of firms whose productivity are above q% of incumbent firms in sector s. It can be shown that $Share_s(q) = \left(\frac{a_s^q}{\bar{a}_s}\right)^{\varepsilon_s - \gamma - 1}$. Further noting that $\frac{q}{100} = \tilde{G}(a_s^q) = 1 - \left(\frac{a_s^q}{\bar{a}_s}\right)^{-\gamma}$, we have $\frac{a_s^q}{\bar{a}_s} = \left(1 - \frac{q}{100}\right)^{-\frac{1}{\gamma}}$ and $Share_s(q) = \left(1 - \frac{q}{100}\right)^{\frac{\gamma+1-\varepsilon_s}{\gamma}}$.

To get an aggregate γ, we use the average of ε_s to eliminate the dependence on s, i.e.,

$$\overline{Share}(q) = \left(1 - \frac{q}{100}\right)^{\frac{\gamma+1-\bar{\varepsilon}}{\gamma}}.$$

We set $\bar{\varepsilon} = \frac{e^{\bar{\mu}}}{e^{\bar{\mu}}-1} = \frac{1.18}{1.18-1} = 6.55$. In China, the top 5% revenue share is 43.53%. Hence the above equation becomes

$$0.4353 = (1 - \frac{95}{100})^{\frac{\gamma+1-6.55}{\gamma}}$$

This implies $\gamma = 7.68$.