Capital Frictions and Misallocation with an S-shaped Production Function

Maitreesh Ghatak * and Dilip Mookherjee †

March 5, 2025

Abstract

We study a 'second-best' model of misallocation resulting from a combination of borrowing constraints (a la Buera, Kaboski and Shin (2021)) and technological nonconvexities (featuring increasing returns over a low range of firm output, followed by decreasing returns thereafter) in a competitive laissez faire economy. The model's predictions are consistent with cross-country differences in misallocation patterns in manufacturing firms documented by Ayerst, Nguyen and Restuccia (2024), thereby constituting an alternative to their explanation based on progressive size-dependent government policies in a first-best model without market frictions. The welfare implications of the two explanations differ markedly: in the second-best model there exist progressive size-dependent policies and wage repression policies that raise welfare, by creating distortions that offset distortions arising from capital market frictions. These results highlight the need to identify the underlying sources of observed misallocation patterns in any given context.

Keywords: misallocation, capital frictions, market failure, second-best, developing countries

1 Introduction

This paper studies a laissez faire competitive economy with capital frictions a la Buera-Kaboski-Shin (BKS, 2021) and extends it to incorporate an S-shaped production function at the firm level. We argue this model can explain stylized facts concerning cross-country variations in manufacturing firm size and productivity distributions (Ayerst, Nguyen and Restuccia (ANR, 2024)):

1. Average firm size is lower, and firm level total factor productivity (TFP) is more dispersed in less developed countries.
2. Larger TFP dispersion in LDCs is driven mostly by greater prevalence of small, low productivity firms.
3. Dispersion of distortions (or 'wedges', measured by average product of labor) is higher in LDCs.[1]
4. Wedges are positively correlated with firm TFP, and this correlation is higher in LDCs.

*London School of Economics; M.Ghatak@lse.ac.uk
†Boston University; dilipm@bu.edu

[1] This is consistent with comparisons of misallocation across specific countries in Hsieh-Klenow (2009).

ANR explain these facts by the presence of progressive size-dependent taxes or government regulations with greater progressivity in LDCs, using a 'first-best' neoclassical model without any market frictions. Progressive government policies stifle output and productivity-enhancement incentives to a greater degree for large firms owned by high ability entrepreneurs, resulting in lower labor demand, lower wages and greater entry of small unproductive firms owned by low ability entrepreneurs. ANR quantify the resulting implications for loss of aggregate TFP using a model with parameters calibrated to French data.

The model we study in this paper features an alternative source of misallocation: capital market frictions and cross-country variation in wealth distributions (where poorer countries have lower average wealth and greater wealth dispersion). Empirical evidence for the role of capital market frictions for misallocation have been provided by a large literature (e.g., including Buera et al (2011), Gopinath et al (2017), Cavalcanti et al (2024)).

We first show such a model is consistent with the stylized facts documented by ANR. With borrowing constraints, entry and firm size depend on an entrepreneur's wealth, besides ability. Holding ability θ constant, an entrepreneur enters if her wealth a exceeds an entry threshold $\hat{a}(\theta)$ and enters with a firm size $S(a, \theta)$ below the first-best level $S^*(\theta)$ if a is smaller than a higher first-best threshold $\bar{a}(\theta)$. Since borrowing constraints limit capital size and other productivity-enhancing investments besides employment, firms owned by constrained entrepreneurs have lower TFP compared to larger ones owned by unconstrained entrepreneurs. Since the wealth distribution is shifted to the left and is more dispersed in a poorer country, it ends up with more small firms with low TFP, and greater size/TFP dispersion (Facts 1 and 2).

If the production function is S-shaped, firms have a U-shaped average cost curve. Hence the average/marginal product of labor (or 'wedge') exhibits an inverse-U with firm size for credit-constrained firms: over a range of 'small' sizes, the wedge increases in firm size upto an 'efficient' size. It decreases thereafter over a 'medium' range. Among large firms with wealthy owners that are not credit-constrained, the wedge is constant. Hence size variations within the small and medium sized group generates dispersion in the wedge, while there is no dispersion within the large category. Poor countries with a preponderance of small and medium sized enterprises can therefore end up with greater dispersion of the wedge, consistent with Fact 3. Finally, the correlation between the wedge and TFP is positive within the small enterprise group, negative within the medium group and zero within the large group. Hence the overall correlation would be positive in poor countries if the fraction of small enterprises is large enough, and would be larger than the correlation in rich countries where most enterprises are of large and medium size, thus accounting for Fact 4.

We then study how the welfare implications differ between the two competing explanations. In the first-best setting, classical welfare theorems apply, and the misallocation generated by progressive size dependent regulations reduce aggregate (utilitarian) welfare. On the other hand, capital market frictions imply the economy operates in a second-best world in the absence of any government interventions. In this setting we show that there generally exists a progressive output dependent policy where taxes paid by large firms are used to subsidize small and medium firms that raises aggregate welfare. This result applies irrespective of the shape of the underlying ability and wealth distributions. It relies on a single sufficient condition which requires that the market friction bites for some but not all ability types. The policy imposes a tax on firms producing above a certain threshold. This creates a distortion causing some high ability types that are not credit-constrained to contract output and bunch just below the threshold where the tax applies. While this distortion is welfare reducing, if the tax is small the welfare loss is second-order because the affected firms are not credit-constrained. On the other hand, the subsidy granted to low output firms relieves their borrowing constraints, resulting in a first order welfare gain which dominates the distortion imposed on the taxed firms. Note that the welfare gain is accompanied by a shift in production from high-TFP large enterprises to low-TFP small and medium enterprises, and increased entry into the latter category, thereby aggravating productive misallocation. Hence the progressive policy induces aggregate TFP and welfare to move in opposite directions in the second-best setting, in contrast with the first-best setting. The result highlights the need to identify the source of misallocation in any given setting.

In order to focus on the essential differences between the first-best and second-best contexts, we abstract from firm dynamics and idiosyncratic TFP shocks.[2] As in the ANR model, we take the ability distribution as given and assume price taking behavior. The key difference is that the distortions in the first-best model are driven by exogenous tax-like wedges which are increasing in the firm's TFP, while in the second-best model they are generated by capital market frictions in conjunction with an exogenously given wealth distribution. Wealth heterogeneity in our model as a source of firm-specific wedges is analogous to i.i.d shocks to firm-specific tax rates in ANR.

Moreover, the wage and interest rate are exogenously given in the baseline version of our model studied in Section 2. This is a more important difference between the two models, as endogeneity of wages plays an important role in the ANR explanation of the stylized facts. In Section 3 we extend our model to incorporate endogenous wages, and show (analogous to Itskhoki and Moll (2019)) in this setting that wage repression policies (implemented via encouraging immigration of low ability agents) have positive first-order welfare effects, in contrast to zero first order effects in the first-best model.

2 Model

2.1 Assumptions

The S-shaped production function of a firm operated by an entrepreneur with ability θ is

$$y = \theta f(S) \tag{1}$$

where

$$\begin{aligned} f(S) &= S_e^{1-\mu} S^{\mu} \quad \text{if} \quad S \leq S_e \\ &= S_e^{1-\delta} S^{\delta} \quad \text{if} \quad S > S_e \end{aligned}$$

where $S_e > 0$ is a technically efficient scale of operation, and $\mu > 1 > \delta > 0$. S denotes the scale at which the firm is operated, which depends on labor employed (n) and investment in capital or other productivity enhancement activities (z) such as worker training, gaining access to better technology or higher quality material inputs:

$$S = z^{\gamma} n^{1-\gamma} \tag{2}$$

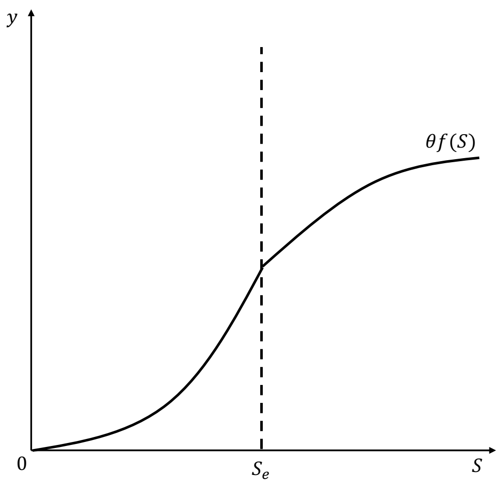
with $\gamma \in (0, 1)$.

The shape of the production function features an initial phase of increasing returns upto the efficient scale S_e (where $\frac{f(S)}{S}$ is maximized), followed by decreasing returns. See Figure 1. A possible interpretation is that all firms have the same production capacity S_e, which is under or over-utilized if S is below or above S_e. If $u \equiv \frac{S}{S_e}$ denotes the utilization rate, $f(S) = S_e u^{\mu}$ if $u \leq 1$ and $= S_e u^{\delta}$ if $u \geq 1$. Note that we get a conventional neoclassical production function with decreasing returns throughout if $\mu = \delta < 1$, and with constant returns throughout if $\mu = \delta = 1$.

Labor is hired at wage rate w and capital (or productivity-enhancing investments) at a rental rate (or price) r which are both exogenously fixed. In Section 3 we extend the model to incorporate endogenous wages.

[2] The dynamics do not play an important role in the ANR model, as it ignores capital accumulation, changes in prices or in the external environment.

FIGURE 1: **S-Shaped Production Function**



Besides variable inputs every firm incurs a fixed cost c to operate. Output price is normalized to unity. We consider a single period, at the beginning of which inputs are procured and paid for. Output and sales are realized at the end of the period.

Agents differ in ability θ and collateralizable wealth a; there is a given joint distribution over these two dimensions of agent heterogeneity represented by conditional cdf $H(a|\theta)$ of wealth of agents of ability θ and marginal cdf $G(\theta)$ over ability). The scale of production is limited by the working capital available to the agent, owing to the borrowing constraint described below. The interest rate on borrowing and lending is the same, so the borrowing constraint constitutes the sole market friction. Let $i \equiv 1 + r$ denote the resulting interest factor.

At the beginning of the period, each agent decides whether to become an entrepreneur, or a worker and earn the given wage w. In the former case, the agent will decide z and n to maximize profits, subject to the borrowing constraint.

2.2 Analysis

The profit of an entrepreneur of ability θ selecting inputs n, z is $\theta f(z^\gamma n^{1-\gamma}) - i(wn + rz + c)$. Given any scale S of operation, n, z will be chosen to minimize $wn + rz$ subject to $S = z^\gamma n^{1-\gamma}$. The solution to this is $n = (1-\gamma)\frac{A}{w}S, z = \gamma\frac{A}{r}S$, resulting in total cost $i(AS + c)$ where $A \equiv [\frac{r}{\gamma}]^\gamma[\frac{w}{1-\gamma}]^{1-\gamma}$. Since w, r are fixed, we can normalize units so that $A = 1$. Then profits equal

$$\pi(S;\theta) - ic \tag{3}$$

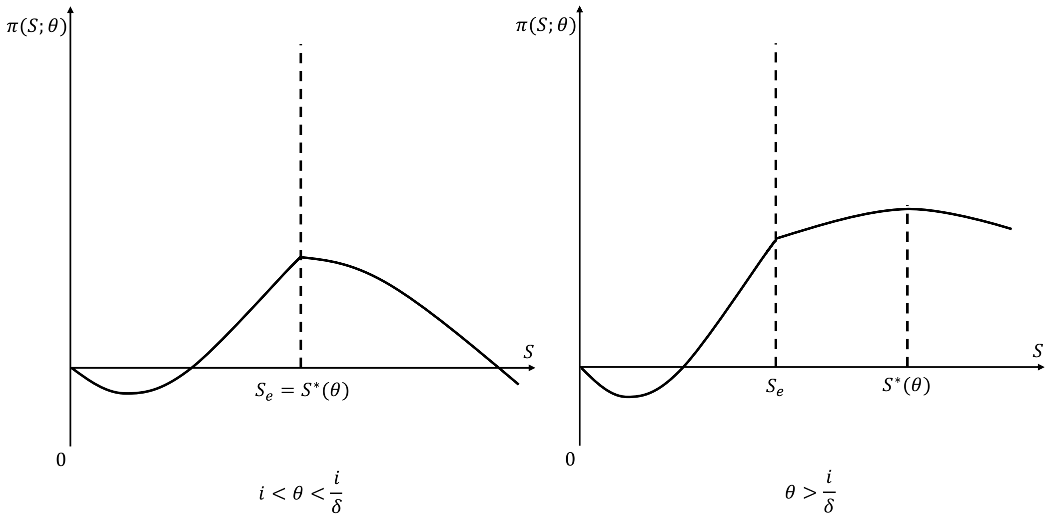
where $\pi(S;\theta)$ denotes operating profits $[\theta f(S) - iS]$, excluding the overhead costs.

2.2.1 First-best Outcomes

In the absence of any borrowing constraint, conditional on operating a firm, an agent of ability θ will select $S^*(\theta)$ to maximize operating profit $\pi(S;\theta)$.

The shape of the operating profit function for two specific values of θ are shown in Figure 2. It is strictly convex over the range $[0, S_e]$, strictly concave over $S > S_e$, with left-hand and right-hand derivatives at S_e equal to $\theta\mu - i, \theta\delta - i$ respectively. Moreover operating profits equal 0 at $S = 0$ and $(\theta - i)S_e$ at $S = S_e$. If $\theta \leq \frac{i}{\mu}$ profits are decreasing and negative at any positive scale. If θ lies between $\frac{i}{\mu}$ and i, profits are negative at S_e and therefore at every positive scale. If θ lies between i and $\frac{i}{\delta}$, as in the left panel of Figure 2, operating profits are initially negative and decreasing at small scales, then rise to $(\theta - i)S_e > 0$ at scale S_e, and fall thereafter. Finally,

FIGURE 2: **Operating Profits (excluding overhead costs)**



if $\theta > \frac{i}{\delta}$, as shown in the right panel of Figure 2, operating profits are initially falling, then rising to positive levels and maximized at $S = S_e[\frac{\delta\theta}{i}]^{\frac{1}{1-\delta}}$ which exceeds S_e, and falls thereafter. Consequently the optimal scale $S^*(\theta)$ for type θ conditional on operation is:

$$\begin{aligned} S^*(\theta) &= 0, \quad \text{if} \quad \theta < i \\ &= S_e, \quad \text{if} \quad i \leq \theta \leq \frac{i}{\delta} \\ &= S_e[\frac{\delta\theta}{i}]^{\frac{1}{1-\delta}}, \quad \text{if} \quad \theta > \frac{i}{\delta} \end{aligned}$$

and the corresponding profits (incorporating overhead costs) are

$$\begin{aligned} \pi^*(\theta) - c &= -ic, \quad \text{if} \quad \theta \leq i \\ &= S_e(\theta - i) - ic, \quad \text{if} \quad i < \theta \leq \frac{i}{\delta} \\ &= S_e[(\delta\theta)^{\frac{\delta}{1-\delta}} - (\delta\theta)^{\frac{1}{1-\delta}}]i^{-\frac{\delta}{1-\delta}} - ic, \quad \text{if} \quad \theta > \frac{i}{\delta} \end{aligned}$$

The agent will become an entrepreneur if and only if $\pi^*(\theta) \geq ic + w$, or $\theta \geq \underline{\theta}^F$ defined by the property that $\pi^*(\underline{\theta}^F) = ic + w$. In what follows we restrict attention to agents of ability at least $\underline{\theta}^F$.

The first-best allocation can be summarized as follows.

If

$$\frac{i}{\delta} \geq i + \frac{ic + w}{S_e} \tag{4}$$

then $\underline{\theta}^F = i + \frac{ic+w}{S_e}$, the first-best allocation is 'partially pooling': agents with $\theta \in [\underline{\theta}^F, \frac{i}{\delta}]$ bunch at S_e while those with $\theta > \frac{i}{\delta}$ choose $S^*(\theta) > S_e$.[3] While if

$$\frac{i}{\delta} < i + \frac{ic + w}{S_e} \tag{5}$$

then $\underline{\theta}^F \in (\frac{i}{\delta}, i + \frac{ic+w}{S_e})$, the allocation is fully separating: all agents that enter choose $S^*(\theta) > S_e$.[4]

[3] $\underline{\theta}^F = i + \frac{ic+w}{S_e}$ in this case because any θ smaller than $i + \frac{ic+w}{S_e}$ will choose S_e if it enters, and at this scale will earn less than w. Meanwhile $\theta = i + \frac{ic+w}{S_e}$ earns w by entering and choosing S_e.

[4] $\underline{\theta}^F \leq i + \frac{ic+w}{S_e}$ because type $\theta = i + \frac{c+w}{S_e}$ can earn at least w by entering and choosing S_e. It can earn strictly

2.2.2 Borrowing Constraint and Second-Best Outcomes

We modify the BKS formulation of the borrowing constraint slightly by requiring all costs $S+c$ to be paid at the beginning of the production period.[5] An agent with assets a would therefore need to borrow if $S+c$ exceeds a. Without loss of generality the borrower borrows $S+c$ and posts his assets as collateral. In the event of a default, the lender can seize the end-of-period value of the borrower's assets ia and a fraction ϕ of profits $\pi(S,\theta)-ic$. The borrower will not default if the default cost $\phi[\pi(S,\theta)-ic]+ia$ exceeds the repayment due $i(S+c)$. This gives rise to the borrowing constraint

$$ia+\phi[\pi(S,\theta)-ic]\geq i[S+c] \tag{6}$$

Consequently, conditional on operating the firm it would choose $S\geq 0$ to maximize $\pi(S,\theta)$ subject to the borrowing constraint (6). Clearly, attention can be restricted to scales $S\in[0,S^*(\theta)]$, since any scale exceeding the first-best level generates less profit than $S^*(\theta)$.

When can the first-best scale $S^*(\theta)$ be financed? This requires $S^*(\theta)$ to satisfy (6), i.e., the agents wealth lies above the threshold $\bar{a}(\theta)$ defined by:

$$\bar{a}(\theta)=\max\{0,S^*(\theta)+c-\frac{\phi}{i}[\pi(S^*(\theta),\theta)-ic]\} \tag{7}$$

and $\bar{a}(\theta)=0$ whenever

$$\phi\geq\phi^*(\theta)\equiv\frac{i(S^*(\theta)+c)}{\pi(S^*(\theta),\theta)-ic} \tag{8}$$

$\phi^*(\theta)$ is a minimum threshold for the enforcement parameter ϕ for *all* agents of ability θ to be able to achieve the first-best, irrespective of their wealth: i.e., $\bar{a}(\theta)=0$.

Lemma 1 *(a) $\phi^*(\theta)$ is decreasing in θ and converges to $\frac{\delta}{1-\delta}$ as $\theta\to\infty$.*

(b) For the first-best to be unattainable for some agents, it is necessary that

$$\phi<\phi^*(\underline{\theta}^F) \tag{9}$$

(c) If

$$\phi\in(\frac{\delta}{1-\delta},\phi^*(\underline{\theta}^F)) \tag{10}$$

there exists $\tilde{\theta}$ such that $\phi^*(\tilde{\theta})=\phi$. Moreover, $\bar{a}(\theta)>0$ and decreasing in θ for all $\theta<\tilde{\theta}$, while $\bar{a}(\theta)=0$ for all $\theta\geq\tilde{\theta}$.

(d) If $\phi<\frac{\delta}{1-\delta}$, $\bar{a}(\theta)>0$ for all θ. It is locally decreasing in θ if $S^*(\theta)=S_e$ locally, and increasing otherwise.

The proof of this and all subsequent Lemmas are provided in the Appendix. From now onwards, for borrowing constraints to matter, we assume enforcement institutions are not strong enough in the sense that (9) holds. If these institutions are of intermediate strength in the sense that (10) holds, part (c) shows that borrowing constraints matter only for agents below some ability level $\tilde{\theta}$. And if enforcement institutions are weak ((d) holds), borrowing constraints bite for every ability θ no matter how large.

more than w by choosing S slightly bigger than S_e. So $\underline{\theta}^F<i+\frac{ic+w}{S_e}$. On the other hand, type $\theta=\frac{i}{\delta}$ cannot earn w, so $\underline{\theta}^F>\frac{i}{\delta}$.

[5] BKS assume instead that workers can be paid at the end of the period, which imply that employment levels are never distorted and the marginal product of labor is equal across all firms. That version would not be able to explain why firm wedges and productivity are correlated or why this correlation may vary across countries.

It remains to analyze optimal decisions of agents with ability $\theta < \tilde{\theta}$ with wealth $a < \bar{a}(\theta)$ who cannot attain the first-best. Conditional on operating, the second-best scale S maximizes $\pi(S, \theta)$ subject to (6). Using the definition of $\pi(S, \theta)$ the borrowing constraint can be rewritten as

$$ia + (1 + \phi)\pi(S, \frac{\phi\theta}{1 + \phi}) - (1 + \phi)ic \geq 0 \tag{11}$$

Since $\pi(S, \frac{\phi\theta}{1+\phi})$ is maximized at $S^*(\frac{\phi\theta}{1+\phi})$, if (11) is not satisfied at $S^*(\frac{\phi\theta}{1+\phi})$ then no S can satisfy it. Hence a necessary condition for an agent of type (a, θ) to be active is that (11) holds at $S = S^*(\frac{\phi\theta}{1+\phi})$, in which case this scale is feasible for the agent. Then $S^*(\frac{\phi\theta}{1+\phi}) < S^*(\theta)$ since $S^*(\theta)$ is not feasible. Over the range $S \in (S^*(\frac{\phi\theta}{1+\phi}), S^*(\theta))$ the function $\pi(S, \frac{\phi\theta}{1+\phi})$ is decreasing. Hence the largest value of S which satisfies the borrowing constraint is $S(a, \theta)$ which solves the equation

$$ia + (1 + \phi)\pi(S, \frac{\phi\theta}{1 + \phi}) - (1 + \phi)ic = 0. \tag{12}$$

It follows that $S(a, \theta)$ is increasing in each argument. And $S(a, \theta)$ converges to $S^*(\theta)$ as a converges to $\bar{a}(\theta)$.

Since we are focusing on agents with $\theta \geq \underline{\theta}^F$ for whom $\pi(S^*(\theta), \theta) \geq ic + w$, and $\pi(S, \theta)$ is increasing over the range of scales where it is nonnegative, there exists a unique $\underline{S}(\theta) \leq S^*(\theta)$ such that $\pi(\underline{S}(\theta), \theta) = ic + w$. This is the minimum scale s at which the agent would want to enter. If $S(a, \theta) \geq \underline{S}(\theta)$ the agent will attain a profit of at least w by entering and selecting scale $S(a, \theta)$. As $\pi(S, \theta)$ is increasing in S over the range $(\underline{S}(\theta), S^*(\theta))$ it will be optimal for the agent to enter and select $S(a, \theta)$ as it is the maximum feasible scale. On the other hand if $S(a, \theta) < \underline{S}(\theta)$ it is optimal for the agent to not enter.

It follows that the agent enters if and only if $S(a, \theta) \geq \underline{S}(\theta)$, and conditional on entering will select $S = S(a, \theta)$. As $S(a, \theta)$ is increasing in a, all agents with ability θ will enter irrespective of wealth if $S(0, \theta) \geq \underline{S}(\theta)$. And if $S(0, \theta) < \underline{S}(\theta)$ the minimum wealth threshold for entry is $\hat{a}(\theta) > 0$ which solves

$$S(a, \theta) = \underline{S}(\theta) \tag{13}$$

This condition states that $\underline{S}(\theta)$ satisfies the equality version of the borrowing constraint:

$$i\hat{a}(\theta) = (1 + \phi)ic - (1 + \phi)\pi(\underline{S}(\theta), \frac{\phi\theta}{1 + \phi)}) \tag{14}$$

More generally, the minimum wealth threshold is defined as follows:

$$\hat{a}(\theta) = \max\{0, \frac{1 + \phi}{i}ic - \frac{1 + \phi}{i}\pi(\underline{S}(\theta), \frac{\phi\theta}{1 + \phi)})\} \tag{15}$$

The second-best allocation can be summarized as follows.

Proposition 2 *For any agent of type (a, θ) with $\theta \geq \underline{\theta}^F$, the second-best allocation is as follows. The agent becomes an entrepreneur if and only if $a \geq \hat{a}(\theta)$ given by (15). Those with $a \in [\hat{a}(\theta), \bar{a}(\theta))$ are credit-constrained and select scale $S(a, \theta)$ (given by (12)) which is locally increasing in a and θ. Those with $a \geq \bar{a}(\theta)$ are unconstrained and select first-best scale $S^*(\theta)$, locally independent of a.*

2.3 Features of the Second-Best Firm Size, Productivity and Wedge Distributions

Some properties of the second-best allocation can be noted (we hereafter refer to scale S as firm size):

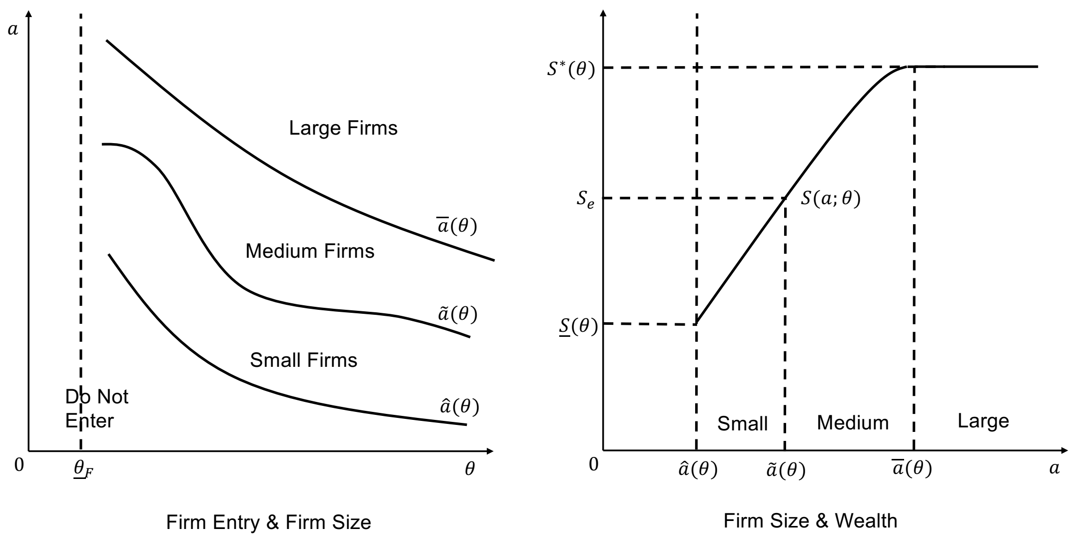
1. Fix ability θ and consider the distribution of firm size generated across agents with this ability and varying wealths. If the wealth distribution conditional on θ has full support, the support of the conditional firm size $S(a, \theta)$ distribution is $[\underline{S}(\theta), S^*(\theta)]$ which does not depend on any other feature of the wealth distribution, i.e., the size distribution is not truncated from below owing to the capital market friction. However, the shape of the conditional size distribution over this support depends on the shape of the wealth distribution. We explore this dependence in more detail below.

2. Allowing θ to also vary, higher θ values correspond to a wider range of firm sizes $[\underline{S}(\theta), S^*(\theta)]$ since $\underline{S}(\theta)$ is decreasing while $S^*(\theta)$ is increasing in θ.

3. Unlike the first-best, firms operated by poor entrepreneurs may select scales below S_e in the second-best. To see this consider any ability $\theta > i + \frac{ic+w}{S_e}$, for whom $\underline{S}(\theta) < S_e$ because operating at scale S_e ensures the agent with ability θ will earn operating profits $(\theta - i)S_e$ exceeding $ic + w$. The range of firm sizes for any such ability will include scales below S_e.

To simplify the exposition in what follows we focus on (i) economies where (5) holds and there is no bunching in the first-best; (ii) ability and wealth are either independent, or positively correlated (in the sense that the conditional wealth distribution at higher ability levels first-order stochastically dominate those at lower levels); (iii) ability levels $\theta > i + \frac{ic+w}{S_e}$, with a minimum scale of operation $\underline{S}(\theta) < S_e$ (this inequality holds since operating at S_e will result in profits that strictly exceed w). Since $S(a, \theta)$ is continuous, and ranges from $\underline{S}(\theta)$ to $S^*(\theta)$ as a ranges from $\hat{a}(\theta)$ to $\bar{a}(\theta)$, we can define an intermediate wealth level $\tilde{a}(\theta)$ where $S(\tilde{a}), \theta) = S_e$.

FIGURE 3: **Second-Best Allocation**



Given Proposition 2, we can classify active firms into three groups:

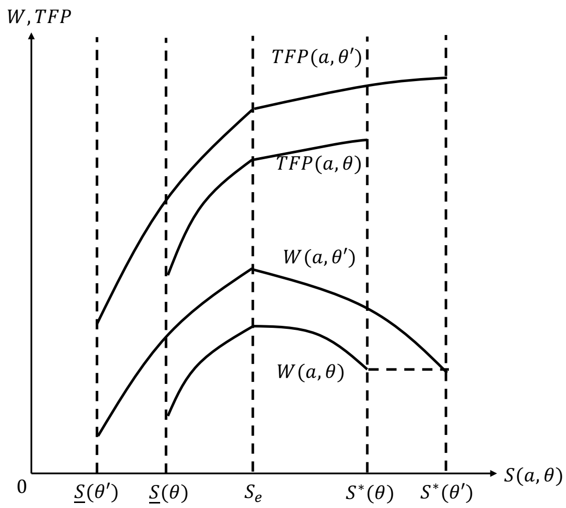
(a) **Small firms:** those with scale $S < S_e$ which are small owing to credit constraints.

(b) **Medium firms:** those with scale $S \in [S_e, S^*(\theta))$ whose owners have intermediate levels of wealth but not large enough to attain the first-best.

(c) **Large firms:** those achieving scale $S^*(\theta)$ owing to their owners wealth exceeding $\bar{a}(\theta)$.

The left panel of Figure 3 shows entry and firm size category outcomes for different combinations of ability and wealth. The right panel shows variations in firm size induced by variations in wealth, holding ability fixed.

Figure 4: **TFP and Wedge Variation with Wealth, holding Ability fixed**



Among **small firms**, output y, TFP $\frac{y}{n^{(1-\gamma)\mu}}$ and wedge $W \equiv \frac{y}{n}$ are:

$$\begin{aligned}\log y &= (1-\mu)\log S_e + \log\theta + \mu\log S(a,\theta)\\ \log TFP &= \gamma\mu\log(\frac{\gamma}{r}) + (1-\mu)\log S_e + \log\theta + \gamma\mu\log S(a,\theta)\\ \log W &= \log C + (1-\mu)\log S_e + \log\theta + (\mu-1)\log S(a,\theta) \end{aligned} \quad (16)$$

where $C \equiv [\frac{\gamma w}{(1-\gamma)r}]^{\gamma\mu}[\frac{1-\gamma}{w}]^{\mu-1}$. Note that $\mu - 1 > 0$ implies the wedge (labor productivity) is increasing in firm size resulting from higher wealth (holding ability fixed), owing to local scale economies over the range of small firm sizes. Hence wealth effects induce co-movement of output, TFP and the wedge.

Since we do not have a closed form solution for the $S(a,\theta)$ function, we compute second-order moments of these distributions using a log-linear approximation $\log S(a,\theta) = \zeta\log a + \nu\log\theta$:

$$\begin{aligned}V(\log y) &= [1+\mu\nu]^2 V(\log\theta) + \mu^2\zeta^2 V(\log a)\\ V(\log TFP) &= [1+\gamma\mu\nu]^2 V(log\theta) + \gamma^2\mu^2\zeta^2 V(\log a)\\ V(\log W) &= [1+(\mu-1)\nu]^2 V(\log\theta) + (\mu-1)^2\zeta^2 V(\log a)\\ COV(\log y, \log TFP) &= [1+\mu\nu][1+\gamma\mu\nu]V(\log\theta) + \gamma\mu^2\zeta^2 V(\log a)\\ COV(\log W, \log TFP) &= [1+\gamma\mu\nu][1+(\mu-1)\nu]V(\log\theta) + \gamma\zeta^2\mu(\mu-1)V(\log a) \end{aligned} \quad (17)$$

It follows that output, TFP and wedge are mutually positively correlated. Output and TFP are positively correlated because increasing ability and wealth induce higher TFP (both directly and indirectly via increased investments in z). And as pointed out above, they increase the wedge (productivity of labor) owing to scale economies ($\mu - 1 > 0$).

(17) also shows that higher wealth dispersion among small firms would generate higher wedge and TFP dispersion within this group.

Among **medium firms**, we obtain analogous expressions with δ replacing μ. Output and TFP continue to be positively correlated, but the sign of the TFP-wedge correlation is now ambiguous (as $\delta - 1 < 0$). Holding ability fixed, an increase in wealth raises investment in productivity enhancement which raises TFP, but lowers W owing to decreasing returns to scale. On the other hand, increasing ability while holding wealth fixed raises both TFP and W. The net effect can go either way. If most of the size dispersion is generated by inequality in wealth rather than ability, i.e.,

$$\frac{V(\log a)}{V(\log\theta)} > \frac{[1+\gamma\delta\nu][1+(\delta-1)\nu]}{\gamma\zeta^2\delta(1-\delta)]} \quad (18)$$

W and TFP would be negatively correlated within the medium firm category.

Finally among **large firms**:

$$\log y = \log\theta + \delta\log S^*(\theta) + (1-\delta)\log S_e \tag{19}$$
$$\log TFP = \gamma\delta\log(\frac{\gamma}{r}) + \log\theta + \gamma\delta\log S^*(\theta) + (1-\delta)\log S_e \tag{20}$$

while the wedge (average/marginal product of labor) is constant, since large firms select first-best scales where the marginal product of labor is equalized. Hence output and TFP are positively correlated, but the wedge is uncorrelated with either.

Figure 4 shows how TFP and W co-vary with firm size as wealth is varied, holding ability fixed at two different levels $\theta' > \theta > i + \frac{ic+w}{S_e}$.

2.4 Explaining the Stylized Facts

These facts concern comparisons between cross-sectional firm size and productivity distributions in developed countries (DCs) and less-developed countries (LDCs). We suppose DCs and LDCs differ only with respect to the wealth distribution: the DC distribution dominates both in the first and second order sense (higher mean, lower dispersion).

Consistent with Fact 1, average firm size would be lower in LDCs, since firm size is increasing in entrepreneur wealth. In particular there would be more small enterprises and fewer large enterprises. Firm level size, output and TFP would be more dispersed in LDCs owing to higher wealth dispersion.

In particular, consistent with Fact 2, higher wealth dispersion is associated with a higher weight in the lower tail of the TFP distribution composed of small enterprises which have lower TFP compared to medium or large enterprises.

Fact 3 states that the dispersion of wedges is larger in LDCs. As shown in Figure 4, holding ability fixed the wedge is rising in wealth among small enterprises, falling in wealth among medium enterprises, and constant for large enterprises. Hence the wedge-TFP relationship exhibits an inverted-U which eventually flattens out at the top. If most DC firms are large while most LDC firms are small, wedge dispersion in DCs would be smaller.

Finally, wedge and TFP would be positively correlated within the small firm category, while the correlation within the medium category could be negative and zero among large firms. Hence the estimated elasticity of wedge with respect to TFP could be positive and large in LDCs, and substantially smaller in DCs, consistent with Fact 4.

2.5 Welfare Implications of Size Dependent Policies

Total welfare in this economy equals aggregate income:

$$\begin{aligned} W &= wG(\underline{\theta}^F) + \int_{\underline{\theta}^F}^{\bar{\theta}} [wH(\hat{a}(\theta)|\theta) \\ &+ \int_{\hat{a}(\theta)}^{\bar{a}(\theta)} \{\pi(S(a,\theta),\theta) - ic\} dH(a|\theta) \\ &+ (1 - H(\bar{a}(\theta))\{\pi(S^*(\theta),\theta) - ic\}] dG(\theta) \end{aligned} \tag{21}$$

where $\bar{\theta}$ denotes the upper bound of the ability distribution (assumed larger than $\underline{\theta}^F$ for the problem to be interesting). The first line represents wage earnings of workers; the second line the profits of constrained entrepreneurs E^c and the third line the profits of unconstrained entrepreneurs E^u. Compared to the first-best, welfare is lower for those with ability above $\underline{\theta}^F$ and (a) wealth below $\hat{a}(\theta)$, who are workers earning w instead of becoming an entrepreneur and earning profit $\pi(S^*(\theta)) - ic$; (b) those with intermediate wealth between $\hat{a}(\theta)$ and $\bar{a}(\theta)$ who are entrepreneurs but earn less than first-best profit owing to a suboptimal firm size. Total output in the economy is lower as a result of these extensive and intensive margins of undercapitalization. Moreover, factors are misallocated between those in E^c and E^u, as factor marginal products vary between entrepreneurs in E^c and E^u, and also between entrepreneurs of varying wealth within E^c.

The model used by Hsieh-Klenow (2009) to derive their misallocation measure (dispersion of marginal products across active firms) is based on the assumption of a fixed aggregate stock of factors and a fixed number of active firms. It does not incorporate welfare losses arising from undercapitalization on either intensive or extensive margins. We now examine welfare effects of some government policies which throw light on the relative magnitude of losses arising due to undercapitalization and productivity dispersion respectively. In particular we show that the capital friction model can rationalize the presence of progressive size-dependent policies, which tax large firms while subsidizing small ones. Welfare increases owing to first order effects of reduced undercapitalization of capital constrained entrepreneurs, which dominate second-order losses associated with increased misallocation of the kind that Hsieh-Klenow (2009) and ARN focus on (e.g, contraction in the size of large firms operated by high ability entrepreneurs, and entry of small firms operated by low ability entrepreneurs). The wedges created by the policy for credit-constrained entrepreneurs thus more than offset the wedges created by the credit constraints. So if the true underlying model is a second-best one with capital frictions, it can justify the presence of progressive size-dependent policies. At the same time, researchers studying the data through the lens of a first-best model would be inclined to attribute the welfare loss to the existence of such policies rather than the underlying market friction.

We consider size-dependent policies of the following form: firms with size (measured by output) exceeding some threshold q^* are required to pay a tax t, while those producing below q^* receive a subsidy $s(t)$. The function $s(t)$ is determined by a budget balance constraint elaborated further below.

The only assumption needed to show that such policies increase welfare is that enforcement institutions are of intermediate strength. As Lemma 1 shows, in this case credit constraints bind only for agents of ability below some threshold $\tilde{\theta}$. The largest firms in the economy consist of those operated by agents with ability θ above $\tilde{\theta}$, none of whom are credit constrained. The progressive policy imposes the tax on firms that produce output larger than $S^*(\tilde{\theta})$, the output produced by agents of ability $\tilde{\theta}$ in the laissez faire outcome. These taxes are used to finance the subsidy for all firms that produce smaller outputs.

Proposition 3 *Suppose enforcement institutions are of intermediate strength, in the sense that (10) holds. There exists a firm size threshold q^* and a policy which imposes a tax t on firms producing more than q^*, and a corresponding subsidy $s(t)$ for all firms with output not exceeding q^*, which (a) balances the government budget and (b) generates higher welfare compared to the second-best laissez faire outcome.*

We start by describing how production and entry decisions are affected by the policy. Let $q^F(\theta) \equiv \theta f(S^*(\theta))$ denote the first-best output for ability θ. Set $q^* = q^F(\tilde{\theta})$. It follows that under laissez faire, firms operated by agents with ability at least $\tilde{\theta}$ produce at least q^* while all other firms produce less than q^*.

Given a tax t on output above q^* and subsidy s for output below q^*, an agent with ability θ at least $\tilde{\theta}$ will either select output q^* or $q^F(\theta) \equiv \theta f(S^*(\theta))$. This is because q^* generates higher profit for such an agent than any other output below q^*, while $q^F(\theta)$ maximizes profit over the range (q^*, ∞). The profit difference between these two options equals

$$d(\theta) \equiv \pi(S^*(\theta), \theta) - \pi(\tilde{S}(\theta), \theta) \tag{22}$$

where $\tilde{S}(\theta)$ is the scale at which an agent of ability θ produces q^*:

$$\theta f(\tilde{S}(\theta)) = q^*. \tag{23}$$

Note that $\tilde{S}(\theta) < S^*(\theta)$ if $\theta > \tilde{\theta}$, and $\tilde{S}'(\theta) < 0$.

Evidently $d(\tilde{\theta}) = 0$ and

$$d'(\theta) = [\pi_\theta(S^*(\theta), \theta) - \pi_\theta(\tilde{S}(\theta), \theta)] - \pi_S(\tilde{S}(\theta), \theta)\tilde{S}'(\theta) \tag{24}$$

which is positive for any $\theta > \tilde{\theta}$ because $\tilde{S}(\theta) < S^*(\theta)$, $\pi_S(\tilde{S}(\theta), \theta) > \pi_S(S^*(\theta), \theta) = i$, $\tilde{S}'(\theta) < 0$ and $\pi_{S\theta}(S, \theta) > 0$. Note also that

$$d'(\tilde{\theta}) = 0 \tag{25}$$

because $S^*(\tilde{\theta}) = \tilde{S}(\tilde{\theta})$ and $\pi_S(\tilde{S}(\tilde{\theta}), \tilde{\theta}) = \pi_S(S^*(\tilde{\theta}), \tilde{\theta}) = i$.

These properties imply that given any $\nu > 0$ we can use the Implicit Function Theorem to define $\epsilon(\nu) > 0$ as follows:

$$d(\tilde{\theta} + \epsilon(\nu)) = \nu \tag{26}$$

$\epsilon(\nu)$ is a smooth increasing function with slope

$$\epsilon'(\nu) = \frac{1}{d'(\tilde{\theta} + \epsilon(\nu))} \tag{27}$$

Moreover $\epsilon(\nu) \to 0$ and $\epsilon'(\nu) \to \infty$ as ν approaches zero. For ν close enough to 0 it follows that $\tilde{\theta} + \epsilon(\nu) < \bar{\theta}$.

These results imply that given a tax t for producing above and subsidy s for producing below q^*, the net disincentive ν for producing above q^* equals $s + t$. It follows that the policy will induce the following reactions from agents of ability at least $\tilde{\theta}$ (where we break ties for boundary types arbitrarily, without loss of generality):

(i) *No Size Effect:* Those with $\theta \in (\tilde{\theta} + \epsilon(s + t), \bar{\theta}]$ produce $q^F(\theta)$ as before and pay the tax t;

(ii) *Contraction:* Those with $\theta \in [\tilde{\theta}, \tilde{\theta} + \epsilon(s + t)]$ produce q^* instead of $q^F(\theta)$ and receive subsidy s.

While production decisions are unaffected for group (i), they fall among those in group (ii) all of whom bunch at the threshold output q^*. The latter effect is the principal welfare-reducing distortion or 'wedge' created by the policy.

Next consider agents with ability below $\tilde{\theta}$. Conditional on becoming an entrepreneur their production decisions are affected as follows. Since the policy disincentivizes producing above q^* which these agents did not want to do even under laissez faire, it follows they will all continue to produce below q^*. And since the subsidy does not vary with the quantity produced, the only way the policy affects their production is by altering their borrowing limits. Effectively, their assets increase by s (conditional on becoming an entrepreneur). So the borrowing limit of an agent of ability θ and assets a increases to $S(a + s, \theta)$, and they select a size of $S(a + s, \theta) = \min\{S(a + s, \theta), S^*(\theta)\}$. Apart from the subsidy, their 'pre-tax' profits increase from $[\pi(S(a, \theta), \theta) - ic]$ to $[\pi(S(a + s, \theta), \theta) - ic + s]$ since they were capital constrained under laissez faire. This represents a welfare gain, as the subsidy partially neutralizes the effect of the capital market friction, thus reducing the aggregate 'wedge' for these agents.

Entry decisions are also affected, since the policy enhances profits of low ability agents that become entrepreneurs. Conditional on $\theta < \tilde{\theta}$, the asset threshold for entry falls from $\hat{a}(\theta)$ to $\hat{a}(\theta, s)$ (we abuse notation slightly by using the same notation for this function as in laissez faire, which can now be written as $\hat{a}(\theta, 0)$) where:

$$\pi(S(\hat{a}(\theta, s) + s, \theta), \theta) + s = ic + w \tag{28}$$

Those with assets a slightly below the laissez faire entry threshold $\hat{a}(\theta, 0)$ who would not have entered under laissez faire now enter with a larger size of $S(a + s, \theta)$, allowing them to earn a pre-subsidy profit strictly higher than w. This represents a welfare gain. So at least some of the additional entry is welfare enhancing. However, for those with assets at or slightly above the new threshold $\hat{a}(\theta, s)$ the profits earned consequent on entering are below w what they were earning under laissez faire. The policy thus encourages excessive entry of low ability, low wealth entrepreneurs, representing an additional welfare loss apart from the capital contraction effect (ii) for high ability entrepreneurs. Hence the net welfare effect of the additional entry is ambiguous.

Summarizing the effects of the policy on agents with ability below $\tilde{\theta}$:

(iii) *Increased Entry:* those with assets $a \in [\hat{a}(\theta, s), \hat{a}(\theta, 0))$ enter; these new entrants select size $S(a + s, \theta)$ and receive subsidy s;

(iv) *Incumbent Expansion:* among incumbents with assets $a \in [\hat{a}(\theta, 0), \bar{a}(\theta))$, capital expands from $S(a, \theta)$ to $\min\{S(a + s, \theta), S^*(\theta)\}$ and they receive subsidy s;

(v) *No Size Effect:* incumbents with $a \geq \bar{a}(\theta)$ continue to produce $q^F(\theta)$ with capital $S^*(\theta)$ and receive subsidy s.

The policy balances the government's budget if total taxes paid by group (i) equals the subsidy received by groups (ii)-(v):

$$t[1 - G(\tilde{\theta} + \epsilon(s + t))] = s \int_{\underline{\theta}}^{\tilde{\theta}+\epsilon(s+t)} [1 - H(\hat{a}(\theta, s)|\theta)]dG(\theta) \tag{29}$$

Lemma 4 *There exists a unique $s(t)$ for any $t \geq 0$ satisfying the budget balance condition (29). The function $s(t)$ is smooth, strictly increasing with $s(0) = 0$ and $s^{'}(0) < \infty$.*

Next, define $e(t) \equiv \epsilon(s(t)+t)$, so $(\tilde{\theta}, \tilde{\theta}+e(t))$ is the range of agent abilities that contract the size of their firms by bunching at q^*. This is a smooth, strictly increasing function satisfying $e(0) = 0$ and

$$e^{'}(t) = \epsilon^{'}(s(t) + t)[s^{'}(t) + 1] = \frac{1 + s^{'}(t)}{d^{'}(\tilde{\theta} + \epsilon(s(t) + t))} \tag{30}$$

which goes to ∞ as $t \to 0$ since s is increasing and $d^{'}(\tilde{\theta}) = 0$.

To calculate the change in welfare resulting from the policy $(s(t), t)$, we can ignore the financial transfers associated with direct payments of taxes and subsidies since the budget is balanced by construction. We need to aggregate the change in 'pre-tax' profits of different groups (ii)-(iv), since these do not change for groups (i) and (v).

The contraction of firm size in group (ii) generates a welfare loss of

$$L(t) \equiv \int_{\tilde{\theta}}^{\tilde{\theta}+e(t)} d(\theta)dG(\theta). \tag{31}$$

Despite the steep increase in the production disincentive $e(t)$ generated by the policy for high ability producers by a small tax starting from laissez faire (recall $e'(0) = \infty$), the next Lemma shows the corresponding effect on welfare is second-order.

Lemma 5 $L^{'}(0) = 0$.

Now turn to the welfare effect of increased entry (group (iii)), which equals

$$E(t) \equiv \int_{\underline{\theta}}^{\tilde{\theta}} \int_{\hat{a}(\theta,s(t))}^{\hat{a}(\theta,0)} [\pi(S(a + s(t), \theta), \theta) - ic - w]dH(a|\theta)dG(\theta) \tag{32}$$

Hence

$$E^{'}(t) = s^{'}(t)[\int_{\underline{\theta}}^{\tilde{\theta}} [\pi(S(\hat{a}(\theta, s(t)) + s(t), \theta), \theta) - ic - w]h(\hat{a}(\theta, s(t))|\theta)dG(\theta) + \int_{\underline{\theta}}^{\tilde{\theta}} \int_{\hat{a}(\theta,s(t))}^{\hat{a}(\theta,0)} \frac{\partial\pi(S(a + s(t), \theta), \theta)}{\partial S} \frac{\partial S(a + s(t), \theta)}{\partial a} dH(a|\theta)dG(\theta)] \quad (33)$$

implying $E^{'}(0) = 0$ as $\pi(S(\hat{a}(\theta, 0), \theta), \theta) - ic - w = 0$ at the laissez faire entry threshold $\hat{a}(\theta, 0)$ which implies the first line of (33) is zero, while the second line is zero as the range of integration shrinks to a single point $\hat{a}(\theta, 0)$.

Finally consider the welfare effect of size expansion in group (iv):

$$X(t) = \int_{\underline{\theta}}^{\tilde{\theta}} \int_{\hat{a}(\theta,0)}^{\bar{a}(\theta)} [\pi(S(a + s(t), \theta), \theta) - \pi(S(a, \theta), \theta)]dH(a|\theta)dG(\theta) \quad (34)$$

implying

$$X^{'}(t) = s^{'}(t) \int_{\underline{\theta}}^{\tilde{\theta}} \int_{\hat{a}(\theta,0)}^{\bar{a}(\theta)} \frac{\partial\pi(S(a + s(t), \theta), \theta)}{\partial S} \frac{\partial S(a + s(t), \theta)}{\partial a} dH(a|\theta)dG(\theta) \quad (35)$$

Hence

$$X^{'}(0) = s^{'}(0) \int_{\underline{\theta}}^{\tilde{\theta}} \int_{\hat{a}(\theta,0)}^{\bar{a}(\theta)} \frac{\partial\pi(S(a, \theta), \theta)}{\partial S} \frac{\partial S(a, \theta)}{\partial a} dH(a|\theta)dG(\theta) \quad (36)$$

which is strictly positive owing to the binding borrowing constraint for this group of entrepreneurs under laissez faire which implies $\frac{\partial\pi(S(a,\theta),\theta)}{\partial S} > 0$, and the relaxation of this constraint owing to the subsidy: $\frac{\partial S(a,\theta)}{\partial a} > 0$ for each member of this group.

Starting from laissez faire, a small t will therefore create a first-order welfare gain owing to the relaxation of borrowing constraints of incumbent entrepreneurs below ability $\tilde{\theta}$, while the corresponding effects of increased entry and contraction of unconstrained entrepreneurs of ability above $\tilde{\theta}$ are second order. This completes the proof of Proposition 3.

3 Endogenous Wages and Wage Repression Policies

Itskhoki and Moll (2019) argue that in the presence of financial frictions, *wage repression* policies could be justified at early stages of industrialization, owing to their positive effect on entrepreneurship. We can illustrate this result and underlying mechanism in the context of our static model. We ignore for simplicity the additional issues that arise in a dynamic setting: Itskhoki and Moll argue that such policies could also benefit workers in the long run by stimulating growth of entrepreneurship which raises labor demand and wages at later stages of industrialization.

The model of the previous section can be extended to endogenous wages in a straightforward manner, while continuing to assume a given interest rate. We need to make explicit the role of the wage rate in firm decisions and extend the notation for firm decisions and outcomes accordingly. The firm's operating profit is now

$$\pi(S, \theta, w) = \theta f(S) - iA(w)S \quad (37)$$

where $A(w) \equiv [\frac{r}{\gamma}]^{\gamma}[\frac{w}{1-\gamma}]^{1-\gamma}$. The optimal employment given scale S equals

$$n(S, w) = (1 - \gamma)\frac{A(w)}{w}S \quad (38)$$

First best scale is now $S^*(\theta, w)$ which is the unconstrained maximizer of $\pi(S, \theta, w)$. And the first-best entry threshold is $\underline{\theta}^F(w)$ defined by $\pi(S^*(\underline{\theta}^F(w), w), \underline{\theta}^F(w), w) = ic + w$. Clearly operating profits and first best scale are both decreasing in w, while the entry threshold is increasing in w.

The borrowing constraint is

$$ia + \phi[\pi(S, \theta, w) - ic] \geq i[A(w)S + c] \tag{39}$$

and the maximum scale consistent with this constraint is $S(a, \theta, w)$ given by the solution to the equality version of (39). This limit is decreasing in w. The asset threshold for achieving the first best is $\bar{a}(\theta, w)$ defined by $S(\bar{a}(\theta, w), \theta, w) = S^*(\theta, w)$, and for entry is $\hat{a}(\theta, w)$ defined by $\pi(\hat{a}(\theta, w), \theta, w) = ic + w$, while the scale of the marginal entrant $\underline{S}(\theta, w)$ is defined by $\pi(\underline{S}(\theta, w), \theta, w) = ic + w$. Those with ability θ and assets between $\hat{a}(\theta, w)$ and $\bar{a}(\theta, w)$ are credit constrained, with profit $\hat{\pi}(a, \theta, w) \equiv \pi(S(a, \theta, w), \theta, w)$ strictly increasing in a, i.e.,

$$\theta f'(S(a, \theta, w)) > iA(w). \tag{40}$$

For unconstrained entrepreneurs with $a > \bar{a}(\theta, w)$, profit equals $\pi^*(\theta, w) \equiv \pi(S^*(\theta, w), \theta, w)$ and is locally independent of a:

$$\theta f'(S^*(\theta, w)) = iA(w). \tag{41}$$

We are now in a position to explain how the wage is determined. The aggregate demand for labor at wage w equals

$$\begin{aligned} N^d(w) \equiv & \int_{\underline{\theta}^F(w)}^{\bar{\theta}} \int_{\hat{a}(\theta,w)}^{\bar{a}(\theta,w)} \Big[n(S(a, \theta, w) dH(a|\theta) \\ & + \{1 - H(\bar{a}(\theta, w)|\theta)\} n(S^*(\theta, w), w) \Big] dG(\theta) \end{aligned} \tag{42}$$

while labor supply equals

$$N^s(w) \equiv G(\underline{\theta}^F(w)) + \int_{\underline{\theta}^F(w)}^{\bar{\theta}} H(\hat{a}(\theta, w)|\theta) dG(\theta) \tag{43}$$

It is easily verified that demand is downward sloping while supply is upward sloping in w, so the equilibrium wage rate is uniquely determined by clearing of the labor market:

$$N^d(w) = N^s(w). \tag{44}$$

Aggregate welfare as a function of w is given by

$$\begin{aligned} W(w) &= w \Big[G(\underline{\theta}^F(w)) + \int_{\underline{\theta}^F(w)}^{\bar{\theta}} H(\hat{a}(\theta, w))|\theta) dG(\theta) \Big] \\ &+ \int_{\underline{\theta}^F(w)}^{\bar{\theta}} \int_{\hat{a}(\theta,w)}^{\bar{a}(\theta,w)} \Big[\{\hat{\pi}(a, \theta, w) - ic\} dH(a|\theta) \\ &+ (1 - H(\bar{a}(\theta, w)) \{\pi^*(\theta, w) - ic\} \Big] dG(\theta) \end{aligned} \tag{45}$$

Now suppose the government introduces a policy which results in a small decrease in w. For instance, it could encourage immigration of foreign workers into the country who cannot become entrepreneurs either owing to business regulations or lack of ability, which would induce a rightward shift of the aggregate labor supply curve. The resulting effect on aggregate welfare of the native population is given by $-W'(w)$ where

$$\begin{aligned} W'(w) &= G(\underline{\theta}^F(w)) + \int_{\underline{\theta}^F(w)}^{\bar{\theta}} H(\hat{a}(\theta,w))|\theta)dG(\theta) \\ &\quad - \int_{\underline{\theta}^F(w)}^{\bar{\theta}} \int_{\hat{a}(\theta,w)}^{\bar{a}(\theta,w)} [-\frac{\partial\hat{\pi}(a,\theta,w)}{\partial w} dH(a|\theta) \\ &\quad -[1-H(\bar{a}(\theta,w)]\frac{\partial \pi^*(\theta,w)}{\partial w}]dG(\theta) \end{aligned} \tag{46}$$

In the above expression, variations in endpoints of integration are ignored owing to indifference conditions prevailing at these endpoints.

Note that the first line of the right-hand-side of (46) equals aggregate labor supply. And the third line equals the labor demand of unconstrained entrepreneurs, since the Envelope Theorem implies that $-\frac{\partial\pi^*(\theta,w)}{\partial w} = n(S^*(\theta,w),w)$. We claim that the double integral expression on the second line is strictly greater than the labor demand of constrained entrepreneurs. This is because for $a \in (\hat{a}(\theta,w), \bar{a}(\theta,w))$:

$$\begin{aligned} -\frac{\partial\hat{\pi}(a,\theta,w)}{\partial w} &= -\frac{\partial\pi(S(a,\theta,w),\theta,w)}{\partial S}\frac{\partial S(a,\theta,w)}{\partial w} - \frac{\partial\pi(S(a,\theta,w),\theta,w)}{\partial w} \\ &= -[\theta f'(S(a,\theta,w),\theta,w) - iA(w)]S_w(a,\theta,w) + iA'(w)S(a,\theta,w) \\ &> iA'(w)S(a,\theta,w) \\ &\geq n(S(a,\theta,w),w) \end{aligned} \tag{47}$$

where the strict inequality on the third line follows from (40) and the fact that $S(a,\theta,w)$ is strictly decreasing in w. The weak inequality on the last line follows from $n(S,w) = (1-\gamma)\frac{A(w)}{w}S = A'(w)S$ and $i = (1+r) \geq 1$.

It follows from the labor market clearing condition that aggregate welfare is strictly decreasing in w, implying that the wage repression policy raises aggregate welfare. This owes to the effect of a decrease in wage rate on the profit of constrained entrepreneurs, which exceeds the number of workers they employ owing to the supplementary effect expanding their borrowing limits. Since workers are worse off, the policy is not Pareto improving.

While both wage repression and the progressive size-dependent policy considered in the previous section raise welfare by reducing undercapitalization on the intensive margin, their distributional effects are different. The burden of relieving borrowing constraints of small and medium entrepreneurs falls on the lowest skill and poorest agents in the wage repression policy. In contrast, the burden is borne by the highest skill and income earning agents in the size-dependent policy. An inequality averse planner is therefore likely to favor the size dependent policy over wage repression.

References

Ayerst, S., Nguyen D.M. and Restuccia D. (2024). The micro and macro productivity of nations. NBER Working Paper 32750.

Buera, F. J., Kaboski, J. P., and Shin, Y. (2021). The macroeconomics of microfinance. Review of Economic Studies, 88:126161.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. The Quarterly Journal of Economics, 124(4):1403–1448.

Itskhoki, O. and Moll, B. (2019). Optimal development policies with financial frictions. Econometrica, 87(1):139–173.

Appendix: Proofs of Lemmas

Proof of Lemma 1: (a) Suppose $\bar{a}(\theta) > 0$. If $S^*(\theta) = S_e$, $\phi^*(\theta)$ equals $i(1 + \frac{c}{S_e})[\theta - i - \frac{c}{S_e}]^{-1}$, which is decreasing in θ. And if $S^*(\theta) > S_e$ it equals $i(1 + \frac{c}{S^*(\theta)})[\frac{i}{\delta} - i - \frac{c}{S^*(\theta)}]^{-1}$ since in this case

$$\phi^*(\theta) = \frac{i(1 + \frac{c}{S^*(\theta)})}{\frac{\pi(S^*(\theta),\theta)}{S^*(\theta)} - \frac{ic}{S^*(\theta)}}$$

and $\frac{\pi(S^*(\theta),\theta)}{S^*(\theta)} = \theta[\frac{S_e}{S^*(\theta)}]^{1-\delta} - i = \frac{i}{\delta} - i$. Hence

$$\phi^*(\theta) = \frac{i(1 + \frac{c}{S^*(\theta)})}{\frac{i}{\delta} - i - \frac{ic}{S^*(\theta)}}$$

which is decreasing in θ since $S^*(\theta)$ is increasing in θ. As $S^*(\theta) \to \infty$ as $\theta \to \infty$, $\phi^*(\theta)$ converges to $\frac{\delta}{1-\delta}$.

(b) If $\phi \geq \phi^*(\underline{\theta}^F)$, we have $\bar{a}(\theta) = 0$ for all $\theta \geq \underline{\theta}^F$.

(c) The existence of a unique $\tilde{\theta}$ satisfying $\phi^*(\tilde{\theta}) = \phi$ follows from (a) and (10). Hence $\bar{a}(\theta)$ is positive for all $\theta < \tilde{\theta}$ and zero for all other θ. If $\bar{a}(\theta) > 0$ it equals $-\frac{\phi}{i}[\pi(S^*(\theta),\theta) - ic] + S^*(\theta) + c$. If $S^*(\theta) = S_e$ over an interval of values for θ, it is obvious that $\bar{a}(\theta)$ is decreasing over this interval. If instead $S^*(\theta) > S_e$:

$$\bar{a}'(\theta) = [\frac{\delta}{1-\delta} - \phi]\frac{S_e}{i}(\frac{\delta\theta}{i})^{\frac{\delta}{1-\delta}} < 0.$$

Finally (d) follows from (a) and the arguments above. ■

Proof of Lemma 4: Rewrite (29) as follows:

$$t = s\frac{1}{[1 - G(\tilde{\theta} + \epsilon(s+t))]}\int_{\underline{\theta}}^{\tilde{\theta}+\epsilon(s+t)}[1 - H(\hat{a}(\theta, s)|\theta)]dG(\theta) \tag{48}$$

Since $\epsilon(s+t)$ is increasing in s and $\hat{a}(\theta, s)$ is decreasing in s, the right-hand-side of (48) is strictly increasing in s. It equals 0 at $s = 0$ and goes to ∞ as $s \to \infty$. By the Implicit Function Theorem, there exists a smooth function $s(t)$ satisfying

$$t = s(t)\frac{1}{[1 - G(\tilde{\theta} + \epsilon(s(t)+t))]}\int_{\underline{\theta}}^{\tilde{\theta}+\epsilon(s(t)+t)}[1 - H(\hat{a}(\theta, s(t))|\theta)]dG(\theta) \tag{49}$$

and $s(0) = 0$. Moreover, differentiating both sides of (49) with respect to t:

$$1 \geq s^{'}(t)\frac{1}{[1 - G(\tilde{\theta} + \epsilon(s(t)+t))]}\int_{\underline{\theta}}^{\tilde{\theta}+\epsilon(s(t)+t)}[1 - H(\hat{a}(\theta, s(t))|\theta)]dG(\theta) \tag{50}$$

because the other dropped terms involving $s^{'}(t)$ in the derivative of the RHS of (49) are all non-negative. We thus obtain an upper bound to the slope of s:

$$s^{'}(t) \leq \frac{1}{\int_{\underline{\theta}}^{\tilde{\theta}+\epsilon(s(t)+t)}[1 - H(\hat{a}(\theta, s(t))|\theta)]dG(\theta)}[1 - G(\tilde{\theta} + \epsilon(s(t)+t))] \tag{51}$$

As $t \to 0$, the RHS of (51) converges to

$$\frac{1 - G(\tilde{\theta})}{\int_{\underline{\theta}}^{\tilde{\theta}}[1 - H(\hat{a}(\theta, 0)|\theta)]dG(\theta)} < \infty$$

completing the proof of Lemma 4.

Proof of Lemma 5: Differentiating (31) with respect to t:

$$L^{'}(t) = e^{'}(t)d(\tilde{\theta} + e(t))g(\tilde{\theta} + e(t)) = [1 + s^{'}(t)][\frac{d(\tilde{\theta} + e(t))}{d^{'}(\tilde{\theta} + e(t))}]g(\tilde{\theta} + e(t)). \tag{52}$$

Since $d(\tilde{\theta}) = d^{'}(\tilde{\theta}) = 0$, L'Hopital's rule implies

$$\lim_{t \to 0} \frac{d(\tilde{\theta} + e(t))}{d'(\tilde{\theta} + e(t))} = \lim_{t \to 0} \frac{d'(\tilde{\theta} + e(t))}{d''(\tilde{\theta} + e(t))} = \frac{d'(\tilde{\theta})}{d''(\tilde{\theta})}. \tag{53}$$

Differentiating (24):

$$\begin{aligned} d^{''}(\theta) \quad = \quad & \pi_{S\theta}(S^*(\theta), \theta)S^*_{\theta}(\theta) - \pi_{S\theta}(\tilde{S}(\theta), \theta)\tilde{S}^{'}(\theta) \\ & + \pi_{\theta\theta}(S^*(\theta), \theta) - \pi_{\theta\theta}(\tilde{S}(\theta), \theta) \\ & - \pi_{SS}(\tilde{S}(\theta), \theta)[\tilde{S}^{'}(\theta)]^2 - \pi_{S\theta}(\tilde{S}(\theta), \theta)\tilde{S}^{'}(\theta) \\ & - \pi_{S}(\tilde{S}(\theta), \theta)\tilde{S}^{''}(\theta) \end{aligned} \tag{54}$$

Evaluated at $\theta = \tilde{\theta}$ where $S^*(\tilde{\theta}) = \tilde{S}(\tilde{\theta})$, the second and fourth lines equal zero. Since $\pi_{S\theta} > 0$, $\pi_{SS}(\tilde{S}(\tilde{\theta}), \tilde{\theta}) = \pi_{SS}(S^*(\tilde{\theta}), \tilde{\theta}) < 0$ and $\tilde{S}(\theta)$ is decreasing, the first and third lines are positive. Hence $d^{''}(\tilde{\theta}) > 0$. The result follows since $d^{'}(\tilde{\theta}) = 0$, $s^{'}(t)$ converges to $s^{'}(0) < \infty$ and $g(\tilde{\theta} + e(t))$ converges to $g(\tilde{\theta})$ as $t \to 0$.