# Capital Frictions and Misallocation with an S-shaped Production Function

Maitreesh Ghatak [*] and Dilip Mookherjee [†]

August 22, 2024

## 1 Introduction

These notes study a laissez faire competitive economy with capital frictions a la Buera-Kaboski-Shin (BKS, 2021) paper (BKS) which incorporate an S-shaped production function. We argue this model can explain the stylized facts concerning firm size and productivity distributions across developed and less-developed countries (Ayerst, Nguyen and Restuccia (ANR, 2024)):

1. Average firm size is lower, and firm level TFP is more dispersed in LDCs
2. Larger TFP dispersion arises mostly due to prevalence of low productivity firms in LDCs
3. Dispersion of distortions (or 'wedges', measured by average product of labor, closely related to the Hsieh-Klenow misallocation measure) is higher in LDCs (in line with specific country comparisons in Hsieh-Klenow (2009))
4. Wedges are positively correlated with firm TFP and this correlation is higher in LDCs

We then argue that the welfare effects of progressive size-dependent policies are different in this model compared with the ARN model.

To facilitate comparison with the alternative theoretical model used in ANR to explain these facts, we use the same notation for different variables as far as possible. We restrict attention to a static setting, in order to simplify the exposition and comparison between the two models. The ARN model is also essentially static as it ignores capital accumulation, changes in prices or in the external environment. And like their model we focus on price taking behavior. The key difference between the two models is that the distortions in their model are endogenously driven by exogenous tax-like wedges which are increasing in the firm's TFP, while in ours they are endogenously generated by capital market frictions.

Other differences between the two models: (i) Unlike ANR, our model allows agents to choose between occupations (whether to become a worker or entrepreneur). This distinction is inessential, as in both models the outside options of entrepreneurs depend on wages. The simpler version of our model treats the wage rate as fixed, but it can be easily extended to incorporate endogenous wages which play an important role in ANR. (ii) Unlike ARN we abstract from idiosyncratic shocks to TFP, but this is inessential. 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 ARN.

## 2 Model

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

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

where

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

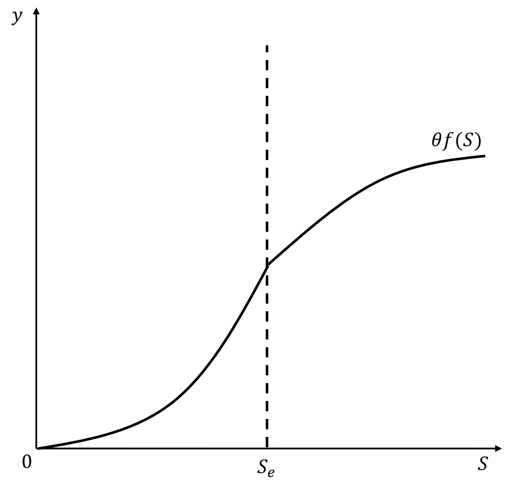
[*]London School of Economics; M.Ghatak@lse.ac.uk
[†]Boston University; dilipm@bu.edu

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

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

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

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




The shape of the production function builds in 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 productivity-enhancing investments at a price $r$ which are given. 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 $\theta$ 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 $\theta$ and marginal cdf $G(\theta)$ over ability). The scale of production is limited by the working capital available to the agent, described by the borrowing constraint described below. While agents face a limit on how much they can borrow, the interest rate on borrowing and lending is the same. Let $i$ denote the resulting interest factor (one plus the interest rate).

At the beginning of the period, each agent decides whether to become an entrepreneur or a worker and earn the given wage $w$.

# 3 Analysis

The profit of an entrepreneur of ability $\theta$ 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 $n = \frac{1-\gamma}{w} S, z = \frac{\gamma}{r} S$, resulting in working capital cost $AS + c$ where $A \equiv [\frac{\gamma}{r}]^{\gamma} [\frac{1-\gamma}{w}]^{1-\gamma}$. Normalize units so that $A = 1$. Then profits equal

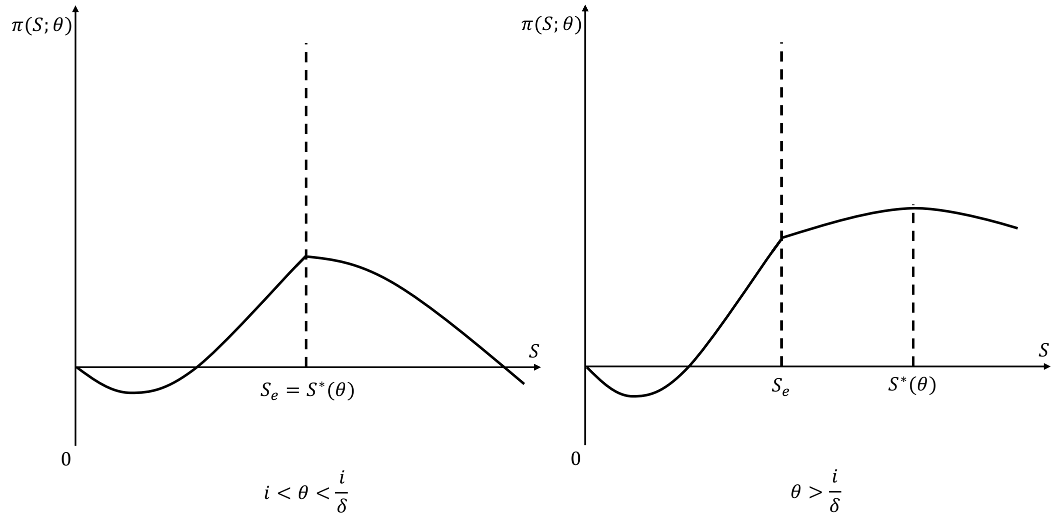
$$\pi(S; \theta) - c \tag{3}$$

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

## 3.1 First-best Outcomes

In the absence of any borrowing constraint, conditional on operating a firm, an agent of ability $\theta$ will select $S^*(\theta)$ to maximize $\pi(S; \theta)$. The shape of the operating profit function for two specific values of $\theta$ are shown in Figure 2.

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




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 $\theta$ lies between $\frac{i}{\mu}$ and $i$, profits are negative at $S_e$ and therefore at every positive scale. If $\theta$ 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, 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 $\theta$ 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

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

The agent will become an entrepreneur if and only if $\pi^*(\theta) \geq c + w$, or $\theta \geq \underline{\theta}^F$ defined by the property that $\pi^*(\underline{\theta}^F) = c + w$. 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

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

then $\underline{\theta}^F = i + \frac{c+w}{S_e}$, 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 $S^*(\theta) > S_e$.[1] While if

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

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

then $\underline{\theta}^F \in (\frac{i}{\delta}, i + \frac{c+w}{S_e})$, the allocation is fully separating: all agents that enter choose $S^*(\theta) > S_e$.[2]

## 3.2 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.[3] 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 $\phi$ of profits $\pi(S, \theta) - c$. The borrower will not default if the default cost $\phi[\pi(S, \theta) - c] + ia$ exceeds the repayment due $i(S + c)$. This gives rise to the borrowing constraint

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

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 (4), i.e., the agents wealth lies above the threshold $\bar{a}(\theta)$ defined by:

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

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

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

$\phi^*(\theta)$ is a minimum threshold for the enforcement parameter $\phi$ for *all* agents of ability $\theta$ 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 $\theta$ 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 $\phi^*(\tilde{\theta}) = \phi$; $\bar{a}(\theta) > 0$ and decreasing in $\theta$ for all $\theta < \tilde{\theta}$; $\bar{a}(\theta) = 0$ for all $\theta \geq \tilde{\theta}$.*

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

**Proof of Lemma 1:** (a) Suppose $\bar{a}(\theta) > 0$. If $S^*(\theta) = S_e$, $\phi^*(\theta)$ equals $i(1 + \frac{c}{S_e})[\theta - i - \frac{c}{S_e}]^{-1}$. And if $S^*(\theta) > S_e$ it equals $i(1 + \frac{c}{S^*(\theta)})[\frac{i}{\delta} - i - \frac{c}{S^*(\theta)}]^{-1}$ where $S^*(\theta)$ is increasing in $\theta$. (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 $\theta$. If $\bar{a}(\theta) > 0$ it equals $-\phi[\pi(S^*(\theta), \theta) - c] + i[S^*(\theta) + c]$. If $S^*(\theta) = S_e$ over an interval of values for $\theta$, it is obvious that $\bar{a}(\theta)$ is decreasing. If instead $S^*(\theta) > S_e$, using the envelope theorem and (4):

$$\bar{a}'(\theta) = \phi\pi_\theta(S^*(\theta), \theta) + iS^{*'}(\theta) = [\frac{\delta}{1-\delta} - \phi]S_e(\frac{\delta\theta}{i})^{\frac{d}{1-\delta}} < 0.$$

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

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

[2] $\underline{\theta}^F \leq i + \frac{c+w}{S_e}$ because type $\theta = i + \frac{c+w}{S_e}$ can earn at least $w$ by entering and choosing $S_e$. It can earn strictly more than $w$ by choosing $S$ slightly bigger than $S_e$. So $\underline{\theta}^F < i + \frac{c+w}{S_e}$. On the other hand, type $\theta = \frac{i}{\delta}$ cannot earn $w$, so $\underline{\theta}^F > \frac{i}{\delta}$.

[3] BKS assume instead that workers can be paid at the end of the period so wage costs, 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.

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 $\theta$ no matter how large.

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 (4). Using the definition of $\pi(S, \theta)$ the borrowing constraint can be rewritten as

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

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 $(a, \theta)$ 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

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

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 $\pi(S^*(\theta), \theta) \geq c + w$, 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 $\pi(\underline{S}(\theta), \theta) = c + w$. 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 $\theta$ 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:

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

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

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

The second-best allocation can be summarized as follows.

**Proposition 2** *For any agent of type $(a, \theta)$ 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 $\theta$. Those with $a \geq \bar{a}(\theta)$ are unconstrained and select first-best scale $S^*(\theta)$, locally independent of $a$.*

# 4 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):

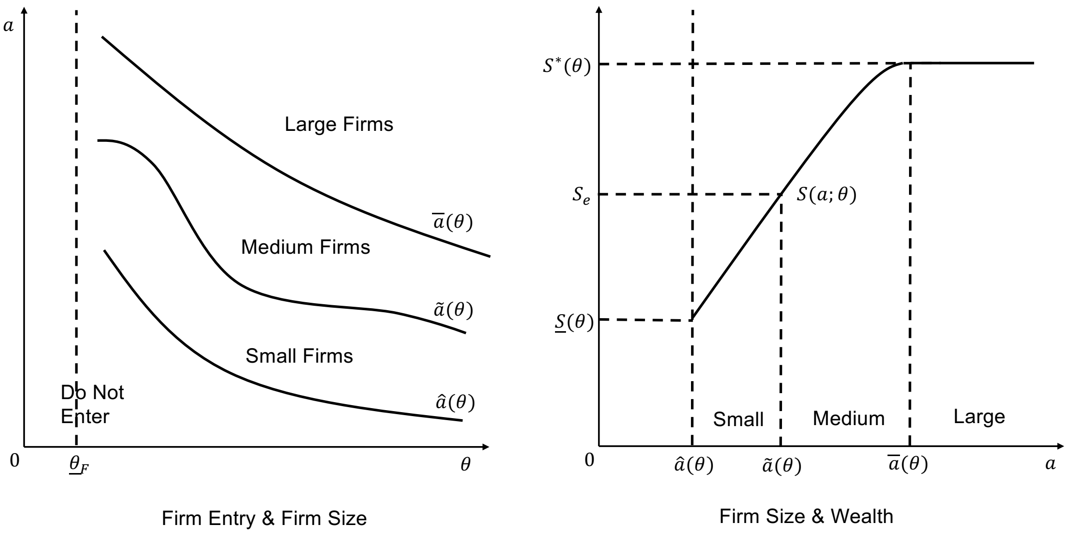
1. Fix ability $\theta$ and consider the distribution of firm size generated across agents with this ability and varying wealths. If the wealth distribution conditional on $\theta$ has full support, the support of the conditional firm size $S(a, \theta)$ distribution is $[\underline{S}(\theta), S^*(\theta)]$ 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 $\theta$ to also vary, higher $\theta$ 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 $\theta$.

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 $\theta > i + \frac{c+w}{S_e}$, for whom $\underline{S}(\theta) < S_e$ because operating at scale $S_e$ ensures the agent with ability $\theta$ will earn operating profits $(\theta - i)S_e$ exceeding $c + w$. 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 (i) (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 $\theta > i + \frac{c+w}{S_e}$), with a minimum scale of operation $\underline{S}(\theta) < S_e$ (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 that bite severely due to low wealth of their owners.

(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.

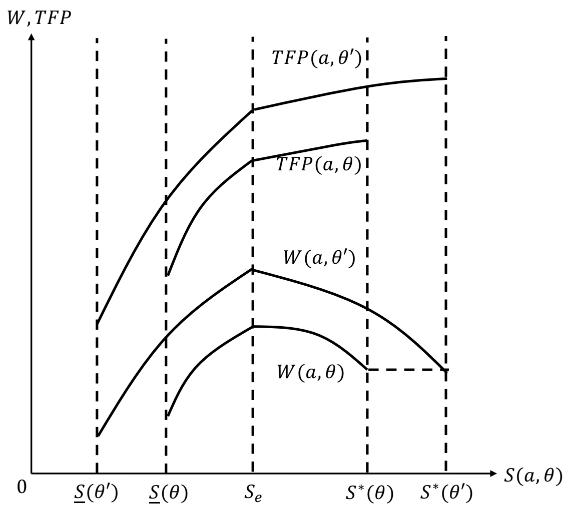
To simplify the exposition we consider the case where

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

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

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.

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




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 $\delta$ replacing $\mu$. 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**:

$$\log y = \log\theta + \delta \log S^*(\theta) \quad (19)$$

$$\log TFP = \gamma\delta \log(\frac{\gamma}{r}) + \log\theta + \gamma\delta \log S^*(\theta) \quad (20)$$

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 $\theta^{'} > \theta > i + \frac{c+w}{S_e}$.

### 4.1 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 **??**, 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.

# 5 Welfare Implications

Total welfare in this economy equals aggregate income:

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

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 $\pi(S^*(\theta)) - c$; (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 $E$, as factor marginal products vary between entrepreneurs in $E^c$ and $E^u$, or of capital compared to the latter group, and also across different entrepreneurs with 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 self-financing 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 $\theta$ 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 $\theta$. 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 $\theta$ 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 $(q^*, \infty)$. 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 $\theta$ 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 $\nu$ approaches zero. For $\nu$ 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 $\nu$ 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 $\theta$ 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 $[\pi(S(a,\theta),\theta) - c]$ to $[\pi(S(a+s,\theta),\theta) - c + s]$ 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:

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

Those with assets $a$ slightly below the laissez faire entry threshold $\hat{a}(\theta, 0)$ who would not have not 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 $S(a+s,\theta)$ 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 $s^{'}(0) < \infty$.*

**Proof of Lemma 5:** Rewrite (30) as follows:

$$t = s\frac{1}{[1 - G(\tilde{\theta} + \epsilon(s+t))]} \int_{\underline{\theta}}^{\tilde{\theta}+\epsilon(s+t)} [1 - H(\hat{a}(\theta,s)|\theta)]dG(\theta) \tag{30}$$

Since $\epsilon(s+t)$ is increasing in $s$ and $\hat{a}(\theta, s)$ is decreasing in $s$, the right-hand-side of (30) is strictly increasing in $s$. It equals 0 at $s = 0$ and goes to $\infty$ as $s \to \infty$. By the Implicit Function Theorem, there exists a smooth function $s(t)$ satisfying

$$t = s(t)\frac{1}{[1 - G(\tilde{\theta} + \epsilon(s(t)+t))]} \int_{\underline{\theta}}^{\tilde{\theta}+\epsilon(s(t)+t)} [1 - H(\hat{a}(\theta,s(t))|\theta)]dG(\theta) \tag{31}$$

and $s(0) = 0$. Moreover, differentiating both sides of (31) with respect to $t$:

$$1 \geq s^{'}(t)\frac{1}{[1 - G(\tilde{\theta} + \epsilon(s(t)+t))]} \int_{\underline{\theta}}^{\tilde{\theta}+\epsilon(s(t)+t)} [1 - H(\hat{a}(\theta,s(t))|\theta)]dG(\theta) \tag{32}$$

because the other dropped terms involving $s^{'}(t)$ in the derivative of the RHS of (31) are all non-negative. We thus obtain an upper bound to the slope of $s$:

$$s^{'}(t) \leq \frac{1}{\int_{\underline{\theta}}^{\tilde{\theta}+\epsilon(s(t)+t)}[1-H(\hat{a}(\theta,s(t))|\theta)]dG(\theta)}[1-G(\tilde{\theta}+\epsilon(s(t)+t))] \tag{33}$$

As $t \to 0$, the RHS of (33) 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 5.

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

$$e^{'}(t) = \epsilon^{'}(s(t)+t)[s^{'}(t)+1] = \frac{1+s^{'}(t)}{d^{'}(\tilde{\theta}+\epsilon(s(t)+t))} \tag{34}$$

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

To calculate the change in aggregate 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

$$L(t) \equiv \int_{\tilde{\theta}}^{\tilde{\theta}+e(t)} d(\theta)dG(\theta). \tag{35}$$

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 $e'(0) = \infty$), we now show the corresponding effect on welfare is second-order.

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

**Proof of Lemma 5:** Differentiating (35) with respect to $t$:

$$L^{'}(t) = e^{'}(t)d(\tilde{\theta}+e(t))g(\tilde{\theta}+e(t)) = [1+s^{'}(t)][\frac{d(\tilde{\theta}+e(t))}{d^{'}(\tilde{\theta}+e(t))}]g(\tilde{\theta}+e(t)). \tag{36}$$

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

$$\lim_{t\to 0}\frac{d(\tilde{\theta}+e(t))}{d^{'}(\tilde{\theta}+e(t))} = \lim_{t\to 0}\frac{d^{'}(\tilde{\theta}+e(t))}{d^{''}(\tilde{\theta}+e(t))} = \frac{d^{'}(\tilde{\theta})}{d^{''}(\tilde{\theta})}. \tag{37}$$

Differentiating (24):

$$\begin{aligned} d^{''}(\theta) \quad = \quad & \pi_{S\theta}(S^*(\theta),\theta)S^*_\theta(\theta) - \pi_{S\theta}(\tilde{S}(\theta),\theta)\tilde{S}^{'}(\theta) \\ & +\pi_{\theta\theta}(S^*(\theta),\theta) - \pi_{\theta\theta}(\tilde{S}(\theta),\theta) \\ & -\pi_{SS}(\tilde{S}(\theta),\theta)[\tilde{S}^{'}(\theta)]^2 - \pi_{S\theta}(\tilde{S}(\theta),\theta)\tilde{S}^{'}(\theta) \\ & -\pi_S(\tilde{S}(\theta),\theta)\tilde{S}^{''}(\theta) \end{aligned} \tag{38}$$

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 $t \to 0$.

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

$$E(t) \equiv \int_{\underline{\theta}}^{\tilde{\theta}}\int_{\hat{a}(\theta,s(t))}^{\hat{a}(\theta,0)}[\pi(S(a+s(t),\theta),\theta) - c - w]dH(a|\theta)dG(\theta) \tag{39}$$

Hence

$$E^{'}(t) = s^{'}(t)[\int_{\underline{\theta}}^{\tilde{\theta}} [\pi(S(\hat{a}(\theta, s(t)) + s(t), \theta), \theta) - c - w]h(\hat{a}(\theta, s(t))|\theta)dG(\theta) + \int_{\underline{\theta}}^{\tilde{\theta}} \int_{\hat{a}(\theta,s(t))}^{\hat{a}(\theta,0)} \frac{\partial \pi(S(a + s(t), \theta), \theta)}{\partial S} \frac{\partial S(a + s(t), \theta)}{\partial a} dH(a|\theta)dG(\theta)] \quad (40)$$

implying $E^{'}(t) = 0$ as $\pi(S(\hat{a}(\theta, 0), \theta), \theta) - c - w = 0$ at the laissez faire entry threshold $(\hat{a}(\theta, 0)$ which implies the first line of (40) 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):

$$X(t) = \int_{\underline{\theta}}^{\tilde{\theta}} \int_{\hat{a}(\theta,0)}^{\bar{a}(\theta)} [\pi(S(a + s(t), \theta), \theta) - \pi(S(a, \theta), \theta)]dH(a|\theta)dG(\theta) \quad (41)$$

implying

$$X^{'}(t) = s^{'}(t) \int_{\underline{\theta}}^{\tilde{\theta}} \int_{\hat{a}(\theta,0)}^{\bar{a}(\theta)} \frac{\partial \pi(S(a + s(t), \theta), \theta)}{\partial k} \frac{\partial S(a + s(t), \theta)}{\partial a} dH(a|\theta)dG(\theta) \quad (42)$$

Hence

$$X^{'}(0) = s^{'}(0) \int_{\underline{\theta}}^{\tilde{\theta}} \int_{\hat{a}(\theta,0)}^{\bar{a}(\theta)} \frac{\partial \pi(S(a, \theta), \theta)}{\partial S} \frac{\partial S(a, \theta)}{\partial a} dH(a|\theta)dG(\theta) \quad (43)$$

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 capital 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.

# 6 Concluding Comments

TBA

# References