Capital Frictions, Nonconvexities and Misallocation

Maitreesh Ghatak [*] and Dilip Mookherjee [†]

August 25, 2025

Abstract

This paper investigates the question whether the source of distortions generating misallocation matters — whether they arise from suboptimal tax-regulation policies in an otherwise perfect market economy, or from market imperfections arising from informational asymmetries or contractual frictions. Ayerst, Nguyen, and Restuccia (2024) have shown that a model of policy-driven distortions is consistent with cross-country patterns in the size-productivity distribution of manufacturing firms. We show that a model of a competitive laissez faire economy with borrowing constraints *á la* Buera, Kaboski and Shin (2011) can also account for the same set of cross-country facts, provided firms' production functions feature increasing returns over an initial range of output, and ability and wealth are not negatively correlated. We then show that the welfare implications of the two models differ markedly. When distortions arise from capital market frictions, there exist progressive size-dependent policies and wage repression policies that raise welfare despite introducing some new welfare reducing distortions, by partially offsetting distortions arising from pre-existing capital market frictions. This holds irrespective of the curvature of the production function or correlation between ability and wealth. These results imply the need to empirically identify underlying source of distortions in any given context before deriving welfare or policy implications.

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

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

1 Introduction

A topic of considerable importance in contemporary macro-development is the extent to which cross-country differences in per capita output or total factor productivity (TFP) can be accounted for by differences in misallocation of *existing* resources, as opposed to technological backwardness or aggregate factor endowments (Banerjee and Duflo (2005), Restuccia and Rogerson (2008, 2017), Hsieh and Klenow (2009, 2014), Buera, Kaboski, and Shin (2011, 2013), Hopenhayn (2014) and Ayerst, Nguyen and Restuccia (2024)). Misallocation is conceptualized as inter-firm variations in distortions or 'wedges' between marginal revenue and costs in varying factor use. Most papers in this literature are agnostic about the specific source of distortions, whether they arise from government taxes and regulations, or from underlying market imperfections. The methodology involves calibration of parameters of explicit dynamic general equilibrium models with agents with heterogeneous ability or innate productivity to fit key moments on the data pertaining to the distribution of firm size and productivity. These calibrated models are subsequently used to show that differences in misallocation account for a substantial proportion of observed variations in per capita output and TFP across countries.

Micro-development economists are also interested in misallocation broadly defined as departures from first-best allocations, arising from market and other institutional imperfections arising from problems of asymmetric information, moral hazard, market incompleteness, weak enforcement of contracts and property rights (Stiglitz 1988, Bardhan and Udry 1998, Pande and Udry 2006). The theoretical micro literature differs from the macro literature by devoting closer attention to distinct positive and normative implications of different sources of frictions, and of heterogeneity of wealth rather than ability. Empirical methodologies involve causal inference of specific sources of imperfection using natural or controlled experiments at the micro level. At the same time less attention is devoted to dynamic general equilibrium spillovers or quantifying implications for aggregate cross-country outcomes.

One of the many conceptual questions raised by the methodological divergence between the macro and micro literatures is the extent to which the specific source of distortions matters. To address this issue, we focus on a set of stylized facts derived by Ayerst, Nguyen and Restuccia (ANR hereafter, 2024) from manufacturing firm data from 28 countries, which they show is consistent with a model of productivity-dependent tax distortions which are more progressive in poorer countries. We then ask whether these facts can also be explained by a model based on credit market imperfections rather than policy-based distortions. We rely on a purely theoretical analysis of a simple static model of credit market imperfections, rather than a quantitative exercise of fitting the model to the firm data.

We focus on credit market imperfections because of the extensive attention devoted to them by both macro and micro-development economists as a major constraint on setting up of firms and their expansion (a recent overview of this literature is provided in Kaboski, 2023). For instance, macro papers studying the impact of financial market imperfections quantitatively include Buera, Kaboski, and Shin (2011, 2013), Gopinath et al (2017), and Cavalcanti et al (2023). Causal evidence regarding the presence of credit rationing in various developing countries with micro-level data using natural or controlled experiments have been provided by de Mel, McKenzie and Woodruff (2008), Banerjee and Duflo (2014), and Breza and Kinnan (2021). There is an earlier literature that used cross-country or within country cross-state variation to study the role of financial markets on growth, such as Rajan and Zingales (1998) and Jayaratne and Strahan (1996).

We first show the ANR stylized facts are also consistent with an alternative model without any taxes but with capital market frictions *á la* Buera, Kaboski and Shin (BKS hereafter, 2011), provided a standard Cobb-Douglas production function is modified to an S-shape which allows for local increasing returns over an initial range of firm scale, followed by decreasing returns thereafter. This model features two-dimensional (ability-cum-wealth) heterogeneity of agents, where wealth and ability are either independent or positively correlated. As in the ANR model, entry, technology and scale decisions are endogenously determined. The cross-country differences in firm size and productivity distributions documented by ANR can be explained by differences in underlying wealth distributions, which are shifted to the left and exhibit greater dispersion in poorer countries.

Having shown that progressive tax distortions and capital market frictions constitute alternative explanations for the same set of facts, we then compare their welfare and policy implications. If misallocation is driven by progressive tax distortions, it is well known that aggregate surplus is maximized by eliminating these taxes. But if is driven by capital market frictions instead, we show that aggregate surplus increases if policies switch from zero taxes to a set of progressive size-dependent taxes and subsidies. This result applies irrespective of the concavity of the underlying production functions, or the specific joint distribution over ability and wealth. Intuitively, these normative results are explained by contrasting patterns of distortions across different size classes of firms under the two different sources. With capital market frictions the distortions are regressive rather than progressive, as poorer owners face more severe credit constraints and are thus prevented from entering or operating their firms at an efficient scale. Consequently there exists a policy which subsidizes small firms and finances these subsidies by levying taxes on firms producing output above some threshold, in which the welfare benefit of offsetting the market-based distortions on small firms outweighs the welfare losses resulting from tax distortions imposed on large firms. Moreover, these aggregate welfare improvements co-exist with symptoms of increased misallocation: increased entry and output of small, less productive firms at the expense of large highly productive firms.

Section 2 provides a more detailed intuitive explanation of the results in the capital friction model. Section 3 provides details of the model and the two main results, for a baseline version where factor prices are exogenously given. In Section 4 we extend the model to incorporate endogenous wages, and show in this setting (analogous to Itskhoki and Moll (2019)) 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. Finally, Section 5 concludes with a brief discussion of implications for future research.

2 Preview of Model and Explanation of Main Results

ANR show the following stylized facts concerning cross-country variations in manufacturing firm size and productivity distributions:

1. Average firm size is lower, and firm level total factor productivity (TFP) is more dispersed in less developed countries (LDCs).
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.
4. Wedges are positively correlated with firm TFP, and this correlation is higher in LDCs.

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). Agents differ in their underlying entrepreneurial ability and (collateralizable) wealth. Similar to the ANR model, entry and technology investments by any given agent are endogenously determined. Provided the production function features increasing returns over an initial range of output, and ability and wealth are either independent or positively correlated, this model turns out to be consistent with the stylized facts documented by ANR for the following reasons. With borrowing constraints, entry and firm size depend both on the entrepreneur's wealth and 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, 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, consider Fact 4 which pertains to the correlation between the wedge (output per worker) and TFP. Both are increasing in the owner's ability, while the effects of wealth variations depend on returns to scale. Among small firms the local increasing returns of the technology implies that the wedge is increasing in the owner's wealth. Hence the wedge-TFP correlation for small firms is positive owing to the variance of their owners' ability and wealth, and positive or zero correlation between ability and wealth. For medium firms the combination of credit constraints and local decreasing returns implies the wedge is locally decreasing in the owner's wealth. Hence the sign of the wedge-TFP correlation within the medium category is ambiguous. For large firms the wedge does not vary with the owners wealth since they are not credit constrained; hence the wedge-TFP correlation is zero for this group. Aggregating across the three size groups, the overall correlation between wedge and TFP would be positive in poor countries if the fraction of small enterprises is large enough in those countries, and would be larger than the correlation in rich countries where most enterprises are of large or 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 returns to scale or how ability and wealth are correlated. It relies on a single sufficient condition that the strength of institutions enforcing loan repayments is neither too high or low, which implies that the market friction bites for some but not all ability types.[1] The welfare improving 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 slightly below the threshold where the tax kicks in. 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.

3 Model

3.1 Assumptions

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

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

[1] We assume enforcement institutions do not vary between poor and rich countries. If they are stronger in richer countries, the contrast in resulting outcomes predicted by our model would likely be intensified further, because the effect of weaker enforcement institutions in raising borrowing constraints is similar to an across-the-board reduction in wealth of all agents.

where

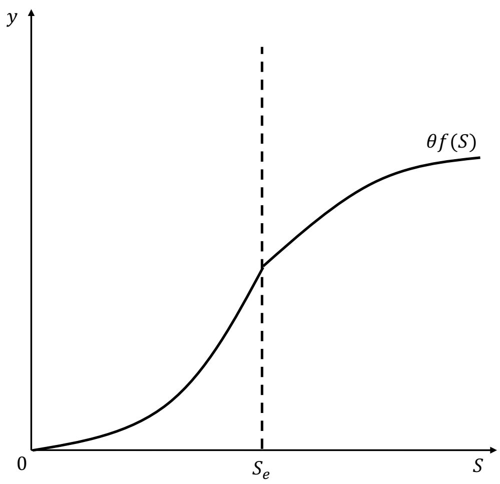
$$
\begin{aligned}
f(S) &= S_e^{1-\mu} S^{\mu} \quad \text{if} \quad S \leq S_e \\
&= S_e^{1-\delta} S^{\delta} \quad \text{if} \quad S > S_e
\end{aligned}
$$

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

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



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.

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 (including fixed costs). Output and sales are realized at the end of the period.

Agents differ in ability θ and collateralizable assets 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 in more detail below. The interest rate on borrowing and lending is the same, so the borrowing constraint constitutes the sole market friction. Let i 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.

3.2 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

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

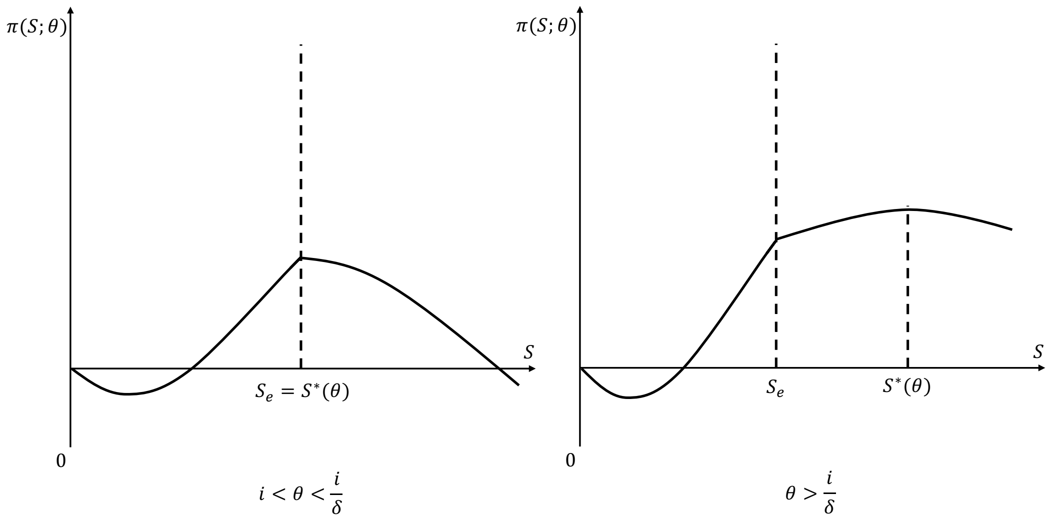
where $\pi(S;\theta)$ denotes operating profits $[\theta f(S) - iS]$, excluding the overhead costs.

3.2.1 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, 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.

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



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 $S^*(\theta) > S_e$.[2] 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 $S^*(\theta) > S_e$.[3]

3.2.2 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.[4] 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., $\bar{a}(\theta) = 0$. It will be shown below

[2] $\underline{\theta}^F = i + \frac{ic+w}{S_e}$ 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.

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

[4] 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.

that

$$\phi^*(\underline{\theta}^F) > \frac{\delta}{1-\delta}. \tag{9}$$

This inequality is useful in classifying the quality of enforcement institutions into different levels as follows.

Lemma 1. *(i) Say enforcement institutions are* **strong** *if $\phi \geq \phi^*(\underline{\theta}^F)$. In this case, the first-best allocation coincides with the second-best allocation for all agents.*

(ii) Say enforcement institutions are **intermediate** *if*

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

In this case, there exists ability level $\tilde{\theta}$ such that all agents with ability at least $\tilde{\theta}$ obtain their first-best allocation irrespective of their wealth, while those with ability below $\tilde{\theta}$ with sufficiently low wealth are credit constrained.

(iii) Say enforcement institutions are **weak** *if*

$$\frac{\delta}{1-\delta} > \phi. \tag{11}$$

In this case, for any ability level, credit constraints bind for agents with sufficiently low wealth.

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, in the sense that (48) holds.

Characterizing the second-best allocation requires us 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

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

Since $\pi(S, \frac{\phi\theta}{1+\phi})$ is maximized at $S^*(\frac{\phi\theta}{1+\phi})$, if (12) 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 (12) 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

$$ia + (1+\phi)\pi(S, \frac{\phi\theta}{1+\phi}) - (1+\phi)ic = 0. \tag{13}$$

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

$$S(a, \theta) = \underline{S}(\theta) \tag{14}$$

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

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

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

$$\hat{a}(\theta) = \max\{0, \frac{1+\phi}{i}ic - \frac{1+\phi}{i}\pi(\underline{S}(\theta), \frac{\phi\theta}{1+\phi})\} \tag{16}$$

The second-best allocation can be summarized as follows.

Proposition 1. *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 (16). Those with $a \in [\hat{a}(\theta), \bar{a}(\theta))$ are credit-constrained and select scale $S(a, \theta)$ (given by (13)) 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.*

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

To simplify the exposition in what follows we focus on economies where:

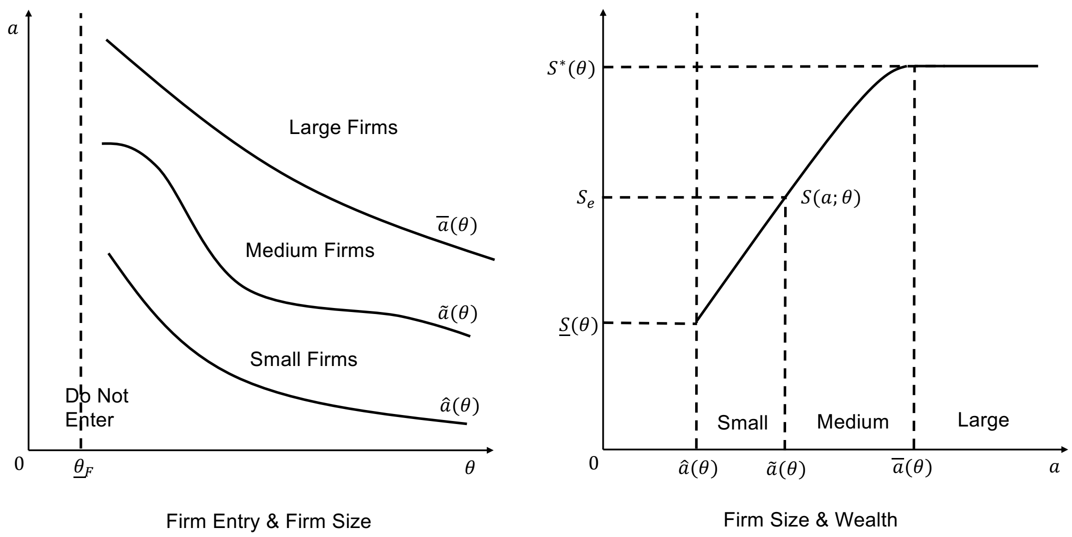
(i) (5) holds and there is no bunching in the first-best;

(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) $\underline{\theta}^F > i + \frac{ic+w}{S_e}$, which ensures that for all relevant ability levels, the minimum scale of operation $\underline{S}(\theta)$ is smaller than S_e.

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 $S(\tilde{a}),\theta) = S_e$. Then agents with wealth below $\tilde{a}(\theta)$ will be underutilizing their capacity owing to credit constraints, and will be subject to local increasing returns to scale. Those with wealth above this threshold will be subject to local decreasing returns.

Figure 3: **Second-Best Allocation**



Consequently we can classify active firms into three groups:

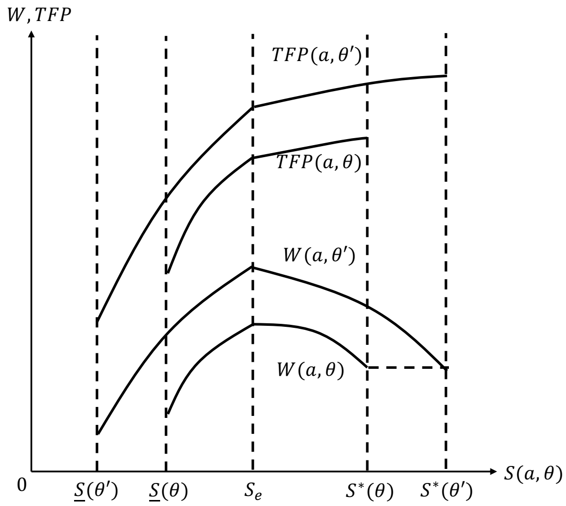
(a) **Small firms:** those with scale $S < S_e$ which operate in the local increasing returns range owing to severely binding credit constraints (owners wealth below $\tilde{a}(\theta)$)

(b) **Medium firms:** those with scale $S \in [S_e, S^*(\theta))$ experiencing local decreasing returns owing to moderately binding credit constraints (owners wealth between $\tilde{a}(\theta)$ and $\bar{a}(\theta)$)

(c) **Large firms:** those achieving first-best 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 at some level above $\underline{\theta}^F$.

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 as follows:

$$\begin{aligned}
\log y &= (1-\mu)\log S_e + \log\theta + \mu\log S(a,\theta) \\
\log TFP &= \gamma\mu\log(\frac{\gamma}{r}) + (1-\mu)\log S_e + \log\theta + \gamma\mu\log S(a,\theta) \\
\log W &= \log C + (1-\mu)\log S_e + \log\theta + (\mu-1)\log S(a,\theta) \qquad (17)
\end{aligned}$$

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 locally increasing returns to scale. Hence among small firms, wealth effects induce positive 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 $\log S(a,\theta) = \zeta \log a + \nu \log \theta$:

$$\begin{aligned}
V(\log y) &= [(1+\mu\nu)^2 + \mu\nu]V(\log\theta) + \mu^2\zeta^2 V(\log a) + \mu\zeta COV(\log\theta, \log a) \\
V(\log TFP) &= [(1+\gamma\mu\nu)^2 + \gamma\mu\nu]V(log\theta) + \gamma^2\mu^2\zeta^2 V(\log a) + \gamma\mu\zeta COV(\log\theta, \log a) \\
V(\log W) &= [\{1+(\mu-1)\nu\}^2 + (\mu-1)\nu]V(\log\theta) + (\mu-1)^2\zeta^2 V(\log a) \\
&\quad +(\mu-1)\zeta COV(\log\theta, \log a) \\
COV(\log y, \log TFP) &= [1+\nu^2+\mu(1+\gamma)(1+\nu)]V(\log\theta) + \gamma\mu^2\zeta^2 V(\log a) + [\gamma\mu^2\zeta\nu \\
&\quad +\mu(1+\gamma)\zeta]COV(\log\theta, \log a) \\
COV(\log W, \log TFP) &= [1+\gamma\mu\nu^2 + (\gamma\mu+\mu-1)\nu]]V(\log\theta) + \gamma\zeta^2\mu(\mu-1)V(\log a) \\
&\quad +[\gamma\mu\zeta\nu + (\gamma\mu+\mu-1)\zeta]COV(\log a, \log\theta) \qquad (18)
\end{aligned}$$

It follows that within small firms, output, TFP and wedge are mutually positively correlated. Output and TFP are positively correlated because increasing ability and wealth induce higher TFP as well as output (both directly and indirectly via increased investments in z); this is further accentuated by positive correlation between ability and wealth. And wedge and TFP are positively correlated because increases in ability and wealth (besides raising TFP) increase firm scale which raises the wedge owing to local increasing returns ($\mu - 1 > 0$).

(18) 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.

Finally among **large firms**:

$$\log y = \log\theta + \delta \log S^*(\theta) + (1-\delta)\log S_e \qquad (19)$$

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

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

3.4 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).[5] Observe that the economy-wide second-order moments of wedge and TFP distributions are weighted averages of the corresponding conditional moments for small, medium and large firm groups, using the proportion of these three groups as weights.

Consistent with Fact 1, average firm size would be lower in LDCs, since firm size is increasing in entrepreneur wealth. 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. If most DC firms are not credit constrained, wedge dispersion would be negligible in DCs. While it would be non-negligible in LDCs if most LDC firms are credit constrained.

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.

Observe also that while the initial range of nonconvexity in the production function in conjunction with positive correlation of ability and wealth constitute sufficient conditions for these results, they are not necessary.

[5] An alternative (or supplementary) mechanism would be weaker enforcement institutions (lower ϕ) in LDCs, which would lower borrowing limits for agents of any given ability and wealth, so the effect would be similar to a reduction in the agent's wealth.

3.5 Welfare Implications of Size Dependent Policies

Total welfare in this economy equals aggregate income:

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

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.

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. 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 2. *Suppose enforcement institutions are of intermediate strength, in the sense that (49) 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 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 2. *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 (proven in the Appendix) shows the corresponding effect on welfare is second-order.

Lemma 3. $L^{'}(0) = 0$.

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

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

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):

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

implying

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

Hence

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

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 2.

4 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

$$\pi(S, \theta, w) = \theta f(S) - iA(w)S \tag{37}$$

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

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

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

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

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

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

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

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

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.

5 Concluding Comments

The main purpose of this paper is to highlight the point that there can be multiple explanations of observed patterns of misallocation, but with very different welfare and policy implications. Our results highlight the need for empirical research to identify the underlying sources of distortions. *A priori* we do not lean one way or the other in terms of whether policies or market imperfections constitutes a better fit to the data. However, empirical identification is likely to prove challenging, in the absence of controlled or natural experiments at the economy-wide level. This may require a methodology which harnesses the strength of the macro-development approach in combination with a more classical approach of trying to see which model provides a better fit to the data (Gabaix and Laibson, 2008).

Addressing this issue is likely to enrich both literatures in different ways and move them closer to one another.[6] Our results illustrate the point that when market imperfections matter, the appropriate normative benchmark is a second-best world where per capita output comparisons can diverge from corresponding welfare comparisons. Macro-development models thus ought to pay more attention to incorporating market imperfections in their quantitative models and to welfare implications of these models. The choice of macro models can be informed and disciplined by micro-studies that incorporate causally identified roles of specific sources of distortions. Conversely, the micro-development literature ought to devote more attention to government taxes and regulations as a possible source of distortion besides market imperfections or wealth inequality, to ability rather than wealth as a source of heterogeneity, and to quantification of aggregate implications of distortions using explicit dynamic general equilibrium models.

References

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

Banerjee, Abhijit V., and Esther Duflo (2005). Growth theory through the lens of development economics. Handbook of economic growth 1 (2005): 473-552.

[6] See Buera, Kaboski and Townsend (2023) and Ghatak and Mookherjee (2025) for an extensive discussion of these issues.

——————— (2014). Do firms want to borrow more? Testing credit constraints using a directed lending program. Review of Economic Studies 81(2): 572-607.

Bardhan, Pranab and Christopher Udry (1999). Development Microeconomics. Oxford University Press.

Breza, Emily, and Cynthia Kinnan (2021). Measuring the equilibrium impacts of credit: Evidence from the Indian microfinance crisis. Quarterly Journal of Economics, 136(3): 1447-1497.

Buera, F. J., Kaboski, J. P., and Shin, Y. (2021). The macroeconomics of microfinance. Review of Economic Studies, 88.1:126-161.

Buera, Francisco J., Joseph P. Kaboski, and Robert M. Townsend. From micro to macro development. Journal of Economic Literature 61.2 (2023): 471-503.

Cavalcanti, Tiago, Kaboski, J. P., Martins, B. S., and Santos, C. Financing costs and development. No. IDB-WP-1526. IDB Working Paper Series, 2023.

de Mel, S., McKenzie, D., & Woodruff, C. (2008). Returns to capital in microenterprises: Evidence from a field experiment, Quarterly Journal of Economics, 123(4), 1329–72.

Gabaix, Xavier and David Laibson (2008). The Seven Properties of Good Models, in Andrew Caplin and Andrew Schotter (eds), The Foundations of Positive and Normative Economics: A Hand Book, Oxford Academic.

Ghatak, Maitreesh, and Dilip Mookherjee. Misallocating misallocation? Annual Review of Economics 17 (2025).

Hopenhayn, Hugo A. Firms, misallocation, and aggregate productivity: A review. Annu. Rev. Econ. 6.1 (2014): 735-770.

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.

Jayaratne, J., and Strahan, P. E. (1996). The finance-growth nexus: Evidence from bank branch deregulation. The Quarterly Journal of Economics, 111(3), 639-670.

Kaboski, Joseph P. (2023). Financial frictions, financial market development, and macroeconomic development, Oxford Development Studies, 51:4, 397-416.

Pande R and Udry C. (2006). Institutions and development: a view from below. In: Blundell R, Newey WK, Persson T, eds. Advances in Economics and Econometrics: Theory and

Applications, Ninth World Congress. Econometric Society Monographs. Cambridge University Press, 349-412.

Rajan, Raghu and Zingales, Luigi (1998). Financial dependence and growth, American Economic Review, 88(3), 559–586.

Restuccia, Diego, and Richard Rogerson. The causes and costs of misallocation. Journal of Economic Perspectives 31.3 (2017): 151-174.

Stiglitz, Joseph (1988). Economic organization, information, and development. In Handbook of Development Economics, Vol. 1, ed. H Chenery, T Srinivasan, pp. 93–160. Elsevier

Appendix: Proofs of Lemmas

Proof of Lemma 1: The proof proceeds by establishing the following claims.

(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

$$\phi < \phi^*(\underline{\theta}^F) \tag{48}$$

(c) If

$$\phi \in (\frac{\delta}{1-\delta}, \phi^*(\underline{\theta}^F)) \tag{49}$$

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.

(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 (49). 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 2: Rewrite (29) as follows:

$$t = s\frac{1}{[1 - G(\tilde{\theta} + \epsilon(s+t))]}\int_{\underline{\theta}}^{\tilde{\theta}+\epsilon(s+t)}[1 - H(\hat{a}(\theta, s)|\theta)]dG(\theta) \tag{50}$$

Since $\epsilon(s+t)$ is increasing in s and $\hat{a}(\theta, s)$ is decreasing in s, the right-hand-side of (50) 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

$$t = s(t)\frac{1}{[1 - G(\tilde{\theta} + \epsilon(s(t)+t))]}\int_{\underline{\theta}}^{\tilde{\theta}+\epsilon(s(t)+t)}[1 - H(\hat{a}(\theta, s(t))|\theta)]dG(\theta) \tag{51}$$

and $s(0) = 0$. Moreover, differentiating both sides of (51) with respect to t:

$$1 \geq s^{'}(t)\frac{1}{[1 - G(\tilde{\theta} + \epsilon(s(t)+t))]}\int_{\underline{\theta}}^{\tilde{\theta}+\epsilon(s(t)+t)}[1 - H(\hat{a}(\theta, s(t))|\theta)]dG(\theta) \tag{52}$$

because the other dropped terms involving $s^{'}(t)$ in the derivative of the RHS of (51) are all non-negative. We thus obtain an upper bound to the slope of s:

$$s^{'}(t) \leq \frac{1}{\int_{\underline{\theta}}^{\tilde{\theta}+\epsilon(s(t)+t)}[1 - H(\hat{a}(\theta, s(t))|\theta)]dG(\theta)}[1 - G(\tilde{\theta} + \epsilon(s(t)+t))] \tag{53}$$

As $t \to 0$, the RHS of (53) 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 2.

Proof of Lemma 3: Differentiating (31) with respect to t:

$$L^{'}(t) = e^{'}(t)d(\tilde{\theta}+e(t))g(\tilde{\theta}+e(t)) = [1+s^{'}(t)][\frac{d(\tilde{\theta}+e(t))}{d^{'}(\tilde{\theta}+e(t))}]g(\tilde{\theta}+e(t)). \tag{54}$$

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

$$\lim_{t\to 0}\frac{d(\tilde{\theta}+e(t))}{d^{'}(\tilde{\theta}+e(t))} = \lim_{t\to 0}\frac{d^{'}(\tilde{\theta}+e(t))}{d^{''}(\tilde{\theta}+e(t))} = \frac{d^{'}(\tilde{\theta})}{d^{''}(\tilde{\theta})}. \tag{55}$$

Differentiating (24):

$$\begin{aligned} d^{''}(\theta) &= \pi_{S\theta}(S^*(\theta),\theta)S^*_\theta(\theta) - \pi_{S\theta}(\tilde{S}(\theta),\theta)\tilde{S}^{'}(\theta) \\ &\quad +\pi_{\theta\theta}(S^*(\theta),\theta) - \pi_{\theta\theta}(\tilde{S}(\theta),\theta) \\ &\quad -\pi_{SS}(\tilde{S}(\theta),\theta)[\tilde{S}^{'}(\theta)]^2 - \pi_{S\theta}(\tilde{S}(\theta),\theta)\tilde{S}^{'}(\theta) \\ &\quad -\pi_S(\tilde{S}(\theta),\theta)\tilde{S}^{''}(\theta) \end{aligned} \tag{56}$$

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 $t \to 0$.