The Role of Finance in the Process of Development: Improving Access versus Reducing Frictions[*]

Timothy Besley
LSE and CIFAR

Konrad Burchardi
IIES, Stockholm

Maitreesh Ghatak
LSE

April 7, 2021

Abstract

We highlight alternative channels through which credit access affects development: capital deepening, financial inclusion, mitigating misallocation by channelling capital to higher productivity entrepreneurs, and raising labour demand and wages through a general equilibrium effect, which also affects the selection into the pool of entrepreneurs. We look at three types of financial frictions: financial access, moral hazard due to limited collateral, and imperfect competition in the credit market. We calibrate a general equilibrium model with contracting frictions, where agents differ in entrepreneurial ability and wealth. We find strong complementarities between different types of financial frictions. In particular, while improving financial access helps unconditionally, the marginal gains are more the lesser are the other frictions. We also find that increased capital market competition has two effects on firm capital: it increases capital access, and in general equilibrium it also increases wages, which in turn affects capital usage directly as well as indirectly by endogenous occupational choice.

JEL Classification: E44, G28, O16

Keywords: Financial Inclusion, Entrepreneurship. Employment Creation

1 Introduction

It is now widely accepted that improving the efficiency of capital allocation is an important aspect of economic development (see, for example, Hsieh and Klenow, 2009, and Restuccia and Rogerson, 2008). Yet the mechanisms that drive misallocation of resources in developing countries are still much debated. One focus is on explanations for capital misallocation, which emphasize the extensive margin with a focus on the economic costs of having large "unbanked" populations with limited access to borrowing and saving opportunities (World Bank, 2014). The second focuses on the intensive margin reducing the "agency costs" by improving the effectiveness of the contracting environment (see, for example, Acemoglu and Zillibotti, 1999). There are many variations of this, including those that focus on the role of competition in the credit market as well as imperfect legal systems that limit the use of assets as collateral (Besley, Burchardi, and Ghatak, 2012).

In order to guide policy priorities in this area, it is useful to understand which of these is quantitatively more important in particular contexts. In this paper we ask the following questions: which forms of alternative financial frictions really matter? For example, should we expand microfinance, should we do asset transfer, should we do bank expansion, should we improve titling and debt collection? What is the inter-relationship between different types of financial frictions - e.g., are they complements or substitutes? Once we allow for general equilibrium effects of better functioning credit markets through labour markets that would raise wages, are these quantitatively important? The indirect effects of some financial intervention programmes may be important, as shown by Bandiera et al (2017) as well as Papp and Imbert (2015).

This paper provides a unified framework for studying the relationship between financial frictions, financial exclusion and the structure of production/level of income. It calibrates a general equilibrium model with contracting frictions and occupational choice where agents differ in entrepreneurial ability and wealth to explore how removing such frictions can raise income levels. A key feature of the framework is that some individuals lack access to financial intermediaries. This is important in practice; the World Bank estimates that about a billion people, the majority of whom are in poor countries, lack access to financial services. Even among those who have access, getting credit is difficult and many reliable on informal borrowing or their own limited wealth.

The key contribution of our paper is highlighting different channels through which credit access affects development and their relative roles. These are: a) capital deepening – how to relax borrowing constraints; b) financial inclusion - making credit available to those who are excluded from the credit; c) misallocation – how better functioning capital markets would allocate capital to higher productivity entrepreneurs; and d) general equilibrium effect – how labour markets are affected, in particular by raising labour demand and wages. In our model this happens through higher labour demand increasing the wage rate, which in turn has a selection effect on who becomes an entrepreneur (in terms of wealth and productivity), and it further affects the functioning of credit markets through the outside option.

We look at three types of financial frictions: financial access – not everyone has access to banking services, moral hazard due to limited wealth/collateral, exacerbated by frictions related to collateralizability of assets (as in de Soto), and imperfect competition in the credit market. Our key findings are as follows.

First, we find that improving financial access helps unconditionally - in the calibrated ver-

sion of our model, we estimate that moving from aurtaky to full-inclusion increases the wage from over 40% to over 80% of the US wage, even with high levels of contracting frictions. This chimes well with claims with the focus on financial inclusion as a key objective of policy. In our framework the gains from inclusion come from three sources: increasing firm size, expanding labour demand, and increasing wages. Since the vast majority of individuals are wage labourers, this spreads the benefits of financial inclusion across the economy. This is consistent with the message of World Bank (2014) and the empirical findings of studies such as Burgess and Pande (2005). This has a direct bearing on which forms of credit market interventions are likely to be most effective - microcredit, credit bureaus, property titling to facilitate collateralization of assets, expansion of bank branches, and mobile banking.

Second, we find strong complementarity between some types of financial frictions. In particular, improving financial access is much more effective the lesser are the other frictions. Therefore financial access is complementary with alleviating these other frictions. For any given level of financial access, starting with bad financial markets (high levels of the both types of financial frictions), alleviating any one friction tends to have a large effect, but alleviating both does not have much more of an additional effect, which suggests these two frictions are not complementary.

Third, we find that increased capital market competition has two effects on firm capital: it increases capital access, and in general equilibrium it also increases wages. Since labour and capital are complementary in our setting the latter effect depresses capital usage. For some entrepreneurs this latter effect more than offsets the positive effect of increased capital access. However, as wages rise, there is stronger selection of entrepreneurs from the pool of those who have higher ability which, in turn, leads the size distribution of firms to shift towards larger employers as capital-deepening takes place in the economy, resulting in higher wages. As the financial market access expands further, we get only a small fraction of the population running their own firms with the vast majority relying on supplying labour, but this is good for the workers as wages are higher.

At the heart of the paper is a simple model of credit market frictions in a general equilibrium setting where lenders compete to serve borrowers. Our characterization of optimal credit contracts transparently highlights a range of interesting economic effects. Our research is in line with the broader research agenda of quantitatively understanding the macro impact of micro frictions, we consider multiple financial frictions and explicitly model the moral hazard problem in credit contract for heterogenous agents: most of the previous research model financial friction as simply collateral constraint (Itskhoki and Moll 2019; Kaboski and Townsend 2011; Midrigan and Xu 2014; Moll 2014; Buera, Kaboski and Shin 2011). Ahrang Lee (2011) considers both the collateral constraint and a fixed cost to using financial services; Dabla-Norris, Ji, Townsend and Unsal (2020) might be the closest to our paper which considers three financial frictions: an entry cost for credit (as measure of financial inclusion), collateral constraint, as well as intermediation inefficiency. And they are able to deliver differential interactions of these frictions in partial equilibrium, which can be substitutes or complements on extensive or intensive margin. We explore different sources of financial frictions and explore their roles in general equilibrium. Also, instead of parameterizing all financial frictions, we write down explicitly the credit contract with moral hazard. This allows us to generate equilibrium default, and heterogenous agents may face different interest rates in general equilibrium (as opposed to the common interest rate among all agents in previous research).

Our approach differs from much of the existing literature which has focused on credit market frictions as *ex post* repayment constraints.[1] In such models, default is absent in equilibrium.[2] In our model, *ex ante* moral hazard leads to defaults in equilibrium, the likelihood of which depends on the extent of collateral a borrower is able to put up. This leads to heterogeneity in default probabilities, and consequently in the interest rates among borrowers who differ in terms of wealth and productivity.

Also, writing down explicitly the credit contract can allow us to explore the role of credit market competition in the effect of financial frictions on development, which is generally lacked in previous research that all assume competitive credit market.

One limitation of our approach is we take a static model. Therefore dynamic incentives and forward-looking savings decisions, which are featured in the most of the related research, are absent in our framework.

The remainder of the paper is organized as follows. In the next section, we discuss the extensive related literature on finance and development. Section three then lays out the theoretical framework. Section four moves from the model to the data and shows how it can be calibrated. Section five develops the results, first on the structure of credit contracts in general equilibrium and second on the effects of extending financial inclusion. Section six concludes.

2 Related Literature

The idea that development of the financial sector has important implications for the economy has a long history with pioneering contributions by Gerschenkron (1962) and Goldsmith (1969). Both put the development banking system at the heart of understanding differences in the trajectories taken by economies. A large body of work has established a strong correlation between measures of financial market development and economic performance at the aggregate level (see, for example, Levine, 2005, Cihak et al, 2013). In parallel, there has also been a theoretical literature on the importance of financial frictions in affecting growth and development including Banerjee and Newman (1993) and Galor and Zeira (1993).[3] Some of this literature focused on heterogeneous entrepreneurial ability (Lloyd-Ellis and Bernhardt, 2000 and Ghatak et al, 2007 being examples), even though now this has become a standard feature in the macrodevelopment literature (Buera et al, 2015). In our model greater entrepreneurial ability has an ambiguous effect on access to credit, conditional on borrowing. On the one hand, it increases the willingness to lend since the marginal productivity of capital is higher. However, we allow for the possibility that larger firms are more costly to manage, and the risk of default can be higher. Nevertheless, improved financial market access enables more able individuals to become entrepreneurs and hire more workers.

This paper is a contribution to the literature which explains income differences through market imperfections. The development accounting literature, such as Caselli (2005), has shown that it is differences in total factor productivity across economies that are key. Our paper is in the spirit of Hsieh and Klenow (2009) who tied this explicitly to capital misallocation.

---

[1]For example, Buera et al (2017) allow the possibility that borrowers may renege on their debt and keep a fraction of the capital, and the only punishment they face is their financial assets deposited with intermediaries forfeited as a result. Such models also tend to imply that all borrowers face the same interest rate.

[2]Models of moral hazard which include equilibrium default are Paulson et al (2006) and Karaivanov and Townsend (2014).

[3]Reviews of the literature can be found in Banerjee and Duflo (2005) and Matsuyama (2007).

This links to older and long-standing debates in the development economics literature on how contracting frictions and imperfect markets matter for under-development. For example, authors such as Bardhan (1984) and Stiglitz (1988) have highlighted a range of such frictions but without providing an approach to assess their implications quantitatively.

There is also an extensive theoretical and empirical literature on how financial arrangements affect households and businesses, particularly how market frictions due to transactions costs and informational constraints may lead to borrowing constraints, and possibly, to poverty traps (see, for example, Banerjee and Duflo 2010, Ghatak, 2015, Karlan and Zinman, 2009, and Townsend and Ueda, 2006).

A number of papers relate financial frictions to aggregate economic performance in ways that combine theory, data and calibration methods. In Jeong and Townsend (2007), there is a modern and subsistence economy with agents differing in wealth and talent. There is a fixed cost of setting up a firm and some agents, as in this paper, lack access to credit markets. They calibrate the model to Thai data showing that credit access is an important factor in explaining TFP dynamics. Buera et al (2011) also study the aggregate implications of credit market access emphasizing that there may be differences between manufacturing and services. They also introduce a non-convexity due to an entry cost. They model the financial friction as due to imperfect enforcement which limits the amount of capital that an entrepreneur can use. After calibrating the model to U.S. data, they find that the variation in financial frictions which they explore can bring down output per worker to less than half of the perfect-credit benchmark. Moll (2014) builds on these approaches and explores the implications of productivity shocks which lead to inefficient capital allocation which can persist in the long-run. This provides a link to research which has looked at the macroeconomic effects of microeconomic distortions such as Bartelsman, Haltiwanger, and Scarpetta (2013), Hsieh and Klenow (2009) and Restuccia and Rogerson (2008). In general, these models do not generate any equilibrium defaults even though they induce capital misallocation.

The literature on financial inclusion has highlighted how having large populations of un-banked populations as a factor constraining development potential in many countries around the world. The Global Financial Inclusion 2014 ("Global Findex") database based on a survey of 150,000 individuals in 148 countries finds sharp differences across countries, showing less use of financial products in poor countries and generally among low income individuals. For example, in developing countries, the top quintile of earners is more than twice as likely to have a bank account than the bottom quintile and the cost of having an account or distance from the nearest branch are frequently cited as the reason (Demirgüç-Kunt and Klapper, 2013).

Correlations between measures of inclusion and development and growth outcomes across countries are strong (see, for example, Sethi and Acharya, 2018, and Sarma and Peis, 2011). But it is hard to interpret reduced-form correlations and issues of reverse causality are rife. Burgess and Pande (2005) exploit a natural experiment due to bank-branching rules in India as a source of exogenous variation and find a significant impact on agricultural wages. The literature linking financial inclusion and economic outcomes is surveyed in Demirgüç-Kunt et al (2017).

Dupas et al (2017) looks at experimental variation in access to banking services in three countries: Uganda, Malawi and Chile. They suggest that there is a puzzlingly low take-up rate of banking services, further underlining the challenge of expanding the outreach of financial services. The extensive literature on the role out of micro-finance to wider populations is

relevant to our findings. However, most of the interventions are at low scale so that the aggregate wage changes that we find here are difficult to study; our analysis emphasises that if there is to be a transformational effect, then structural change will be at its heart.

3   Theory

Our starting point is the standard model of lending under *ex ante* moral hazard and limited liability as in Besley et al (2012).[4] A group of agents who are heterogeneous in two dimensions: entrepreneurial productivity and wealth can choose one of two possible occupations: becoming an entrepreneur or being an employee. If they choose to be entrepreneurs, then they have to decide how much capital and labour to employ.

There are two phases of hiring: managerial labour input is chosen up front and determines the likelihood of creating a successful firm and workers are hired from the labour market only after it is known whether the initial venture has resulted in a successful firm. Capital can come from the entrepreneur's own resources, i.e. their wealth, but can be augmented by borrowing if they have access to financial markets. Lending is risky because some firms are not successful.

Credit markets are subject to two key frictions, as is standard in models of moral hazard with risk neutrality and limited liability.

First, the level of managerial input which determines the likelihood of creating a successful firm is not observed by lenders.

Second, some entrepreneurs lack sufficient wealth to post as collateral and there is limited liability. This rules out the possibility of making entrepreneurs full residual claimants, creating an *ex ante* moral hazard problem. Wealth that can be used as collateral can be limited either because borrowers are poor or due to imperfections in the legal system that limits the collateral value of a given amount of wealth. To capture the latter, we suppose that if a borrower pledges wealth a as collateral to become an entrepreneur, and the firm is not successful, then the bank only recovers a fraction of that collateral.

Lenders design optimal credit contracts subject to information and wealth constraints. Contracts must also respect the entrepreneur's reservation payoff, which is determined endogenously by his outside opportunities whether by borrowing from another lender or working as an employee.

We begin by studying optimal credit contracts which reflect the characteristics of entrepreneurs with a fixed outside option. This illustrates how frictions in the credit market lead to misallocation of capital due to the risk premium that is charged to compensate for the probability of default. We then consider the option of borrowing from a different lender. We capture competition by varying the fraction of the surplus which goes to the borrower as opposed to the lender. A fully competitive credit market is where the entrepreneur gets all of the surplus whereas the opposite is true with competition.

We then introduce a financial inclusion parameter which, following Jeong and Townsend (2007), denotes the fraction of individuals with access to financial markets. Following Townsend (1978), we think of this as reflecting a fixed transaction cost that some agents face (e.g., due to

---

[4]Paulson et al (2006) and Karaivanov and Townsend (2014) estimate models with moral hazard on Thai data with equilibrium default. They do not have managerial labour or focus on endogenous wages. However, they have a rich set of household outcomes with endogenous consumption.

their geographical location or level of knowledge). Entrepreneurs without financial access can only set up firms use their own wealth.

We then study occupational choice for each type of entrepreneur depending on their wealth, productivity and access to financial markets. Entrepreneurs tend to be drawn from among the most wealthy and productive individuals.

Finally, we determine wages endogenously using a standard decreasing-returns Lucas "span of control" model. In general equilibrium, the equilibrium wage and the fraction of agents who become entrepreneurs are jointly determined along with the outside option of agents who borrow to become entrepreneurs. This, in turn, affects the structure of credit contracts.

3.1 Entrepreneurs, Managers, and Workers

The economy is populated by a continuum of agents who are endowed with a unit of time which they supply as labour inelastically regardless of their occupation. All agents are risk neutral and each individual makes a discrete occupational choice, whether to become an entrepreneur and set up a firm, or to become an employee, i.e. work for a firm. Entrepreneurs earn profits from the firm that they own while employees are paid a wage. The wealth that agents have a certain endowment of is denoted by a, which varies across the population.[5]

Entrepreneurs   Entrepreneurs commit all of their labour time to their own firm and are residual claimants on the firm's profit stream. Their ability as entrepreneurs is indexed by θ. Heterogeneous productivity can be interpreted either as entrepreneurial "ability" or having access to a particular production technology.[6] Entrepreneurs hire two kinds of employees: managers who contribute towards the success in the initial start-up phase, and workers who increase output in already successful firms.

We denote the level of managerial input by e and the probability that an entrepreneur creates a successful firm is given by $g(e; \theta) \in [0, 1]$ which is an increasing function of managerial input.[7] We are agnostic about how $g(e; \theta)$ depends on θ. More able entrepreneurs could potentially enhance the productivity of managerial input. However, if high θ entrepreneurs use more complex technologies or spread themselves more thinly over larger firms, then all else equal, this could lower the probability of creating a successful firm. If the start-up is successful and a firm is set up, output is given by $f(k, l; \theta)$ where k is the capital employed and l is wage labour employed.

We make the following regularity assumptions:

Assumption 1 *The following conditions hold for $g(e; \theta)$ and $f(k, l; \theta)$:*

(i) $g(e; \theta)$ is strictly increasing, twice-continuously differentiable, strictly concave for all e with $g(0; \theta) = 0$.

---

[5]The level of wealth is specified in units of labour endowment

[6]If there were a frictionless market for ideas then entrpreneurial ability would no longer matter and ideas would be sold to agents with the highest wealth. Hence, we are assuming that contracting frictions prevent this from happening.

[7]This formulation generalizes the standard agency formulation where success depends only on an entrepreneur's own unobserved effort. Hiring managers to increase e allows the entrepreneur to spread her talent to a wider span of control.

(ii) $f(k, l; \theta)$ *is twice-continuously differentiable and strictly increasing in $k \in \mathbb{R}^+$ and $l \in \mathbb{R}^+$, strictly concave in l, and is increasing in θ with $f_{\theta k} > 0$ and $f_{\theta l} > 0$. Further $f(k, l; \theta) \geq 0$ for all $(k, l) \in \mathbb{R}^+ \times \mathbb{R}^+$ and $f(0, l; \theta) = 0$.*

(iii) $\epsilon(e; \theta) := -\frac{g_{ee}(e; \theta)}{(g_e(e; \theta))^2}$ *is continuous and increasing for all e such that $g(e; \theta) \in [0, 1]$.*

These are more or less standard assumptions which hold in commonly-used models such as Cobb-Douglas and with constant elasticity formulations of the technology which we use in the calibration below. The last part of this assumption guarantees that the level of managerial input increases when the entrepreneur's outside option improves.

The Production Process There are two stages to the production process.

At stage one entrepreneurs negotiate credit contracts with lenders and hire managerial labour. The stochastic nature of firm success which generates the possibility of default. We assume that entrepreneurs submit "business plans" to potential lenders which specify e and which also reveal θ and a to the lender. Lenders understand that, since hiring managerial input is costly and cannot be monitored *ex post*, there is the potential for moral hazard. They will anticipate this when they design contracts.

If the firm is successful, then workers, l, are hired. If the firm fails, capital as well as the managerial labour is wasted and the lender and the managers are not paid. Hence, managers and lenders need to be compensated for this risk. However, no risk is born by workers who are only hired in successful firms.

Employees Agents who choose not to become entrepreneurs are employees. We assume that they are equally productive in managerial task (supplying e) or as wage labour (supply l). A wage labourer earns p_l while a manager earns p_m. As long as they are compensated for the risk of being a manager due not being paid if the firm fails, a risk neutral agent will be indifferent between workers and managers and will also not care which firm they work for. Since the risk is firm specific this means that observable wages for managers will vary by firm productivity, θ.

The Price Vector The output price is p_y. Henceforth, let $\mathbf{p} = (p_y, p_l, p_m)$ denote the price vector. Without loss of generality, and for notational compactness, we allow *all* of the functions that depend on *any* price to be functions of the *entire* price vector even if only some prices are relevant for some specific decisions. In general equilibrium, these prices will be determined endogenously.

3.2 Lenders and Credit Contracts

Wealth and Collateral Consider a slight variation of the model of Besley et al (2013). Let x be the loan size. The total capital invested in the project is $k := x + \psi a$. If the project succeeds (with probability $g(e; \theta)$), the lender receives a gross *payment* of r (which means that the implied net interest rate is $\frac{r}{x} - 1$). If the project fails (with probability $1 - g(e; \theta)$) the lender

captures c. The opportunity cost per unit of capital for the lender is $\gamma \geq 1$ (with $\gamma - 1$ being net rate of interest). Lenders can all access funds at a constant marginal opportunity cost γ.[8]

A fraction ψ of the borrower's assets a can be liquidated at no cost and invested in the project. Any assets that have not been liquidated yield a return γ (the same as the rate of return on liquid assets).[9] Any assets that have not been liquidated, $(1 - \psi)a$, can be pledged as collateral. Given the rate of return γ, the potential collateral value to the lender is $\gamma (1 - \psi) a$. Let τ_1 be the probability the lender is able to seize this form of collateral. Without any frictions, $\tau_1 = 1$. The fact that $\tau_1 < 1$ reflects possible frictions associated with liquidating the entrepreneur's personal assets, giving an expected value of collateral of $\tau_1 \gamma (1 - \psi) a$ from this source.

Suppose also that a fraction δ of the firm's capital can be liquidated in case of failure. Again, these assets can be pledged as collateral to the lender. The potential collateral value from liquidating the capital invested in the firm is $\delta (x + \psi a)$. Let τ_2 be the fraction of the liquidated value of the capital invested in the firm that the bank can seize as collateral, giving an expected value of collateral from this source of $\tau_2 \delta (x + \psi a)$. The borrower's expected payoff from liquidating the firm's assets is $(1 - \tau_2) \delta (x + \psi a)$.

Let $\pi (k; \theta, \mathbf{p})$ denote the conditional profit function given an allocation of capital k which will be defined below.

First-best In the first-best, the allocation decision consists of choosing effort e and capital k to maximize expected total surplus:

$$\max_{e,k} S(e, k; \theta, \mathbf{p}) := g(e; \theta) \pi(k; \theta, \mathbf{p}) + [1 - g(e; \theta)] \delta k - \gamma k - p_m e.$$

How $k = x + \psi a$ is split between self-financing by the borrower (ψa) and borrowing from the lender (x) does not matter, i.e., the first-best total surplus does not depend on ψ.

The first-order conditions with respect to e and k are:

$$g_e(e; \theta) [\pi(k; \theta, \mathbf{p}) - \delta k] = p_m \tag{1}$$
$$g(e; \theta) \pi_k(k; \theta, \mathbf{p}) + [1 - g(e; \theta)] \delta = \gamma. \tag{2}$$

The second order conditions for the existence of a unique maximum require that the functions $g(e; \theta)$ and $\pi(k; \theta, \mathbf{p}) - \delta k$ are not just concave but are "sufficiently" concave.[10] For ex-

---

[8]This could be justified by supposing that this is a small open economy that faces a given international interest rate. Otherwise, we would have to close the model with an endogenous γ which equated the demand and supply of loanable funds.

[9]We assume that illiquid assets earn the same market return γ as liquid capital which is consistent with the fact that they can be liquidated costlessly at any point. In a world where houses can be bought and sold with no transactions costs or risks or value appreciation, the returns on them and the interest rate should be the same.

[10]The second derivatives with respect to e and k are $S_{ee} = g_{ee}(e; \theta) \{\pi(k; \theta, \mathbf{p}) - \delta k\}$ and $S_{kk} = g(e; \theta) \pi_{kk}(k; \theta, \mathbf{p})$ and the cross partial derivative with respect to e and k is $S_{ek} = g_e(e; \theta) \{\pi_k(k; \theta, \mathbf{p}) - \delta\}$. The second-order conditions are: $S_{ee} < 0$, $S_{kk} < 0$ and $S_{ee} S_{kk} > (S_{ek})^2$. The first two conditions are satisfied for any strictly concave function. The third one implies

$$\frac{\{-g_{ee}(e; \theta)\} g(e; \theta)}{\{g_e(e; \theta)\}^2} \frac{\{-\pi_{kk}(k; \theta, \mathbf{p})\} \{\pi(k; \theta, \mathbf{p}) - \delta k\}}{\{\pi_k(k; \theta, \mathbf{p}) - \delta\}^2} > 1.$$

If this holds for all e and k then the objective function is globally concave and therefore the second-order conditions are also sufficient to guarantee the existence of a global optimum.

ample, for $g(e) = \theta e^\alpha$, the condition $\frac{-g''(e)g(e)}{(g'(e))^2} > 1$ is equivalent to $\alpha < \frac{1}{2}$.[11]

Let us denote by e_{FB} and k_{FB} the solution to the pair of equations given by the first-order conditions. Then expected total surplus is $S(e_{FB}, k_{FB}; \theta, \mathbf{p})$. For the first condition to yield an interior solution, we require $\pi(k; \theta, \mathbf{p}) - \delta k > 0$. Notice that this, together with our assumption that $\pi(k; \theta, \mathbf{p})$ is concave, the second first-order condition and $g(e; \theta) < 1$ imply[12]

$$\pi_k(k; \theta, \mathbf{p}) > \gamma > \delta.$$

Any interior solution hence requires that $\gamma > \delta$. As δ is the fraction of the capital that can be salvaged from a failed project, $\delta \leq 1$ and so this condition holds given our assumption $\gamma \geq 1$, which is reasonable as it implies that the net return on capital is non-negative.

Second-best Credit contracts are described by a vector $(x, r, c, \psi a)$ comprising (i) an amount borrowed, x, (ii) an amount to be repaid if the firm is successful, r, (iii) an amount of financial collateral c, (iv) the borrower's equity ψa. For notational simplicity we will use $\mathbf{t} = (x, r, c, \psi a)$ to denote a credit contract.

A lender's expected profit when agreeing to lend to an entrepreneur with collateral c is therefore:
$$\Pi(e, \mathbf{t}; \theta) = g(e; \theta) r + [1 - g(e; \theta)]c - \gamma x. \tag{3}$$

There is a finite set of lenders with whom entrepreneurs can contract. To model competition between lenders, we suppose that there is a Bertrand-style price setting game. Imagine that there are two lenders with identical access to the capital market, γ and the same enforcement technologies. In principle this should lead to borrowers capturing all of the surplus as lenders compete for borrowers until *ex ante* payoffs are zero. However, there are good reasons to doubt that this is a reasonable model and there are likely to be costs of switching between lenders. Rather than being specific about the friction, we capture imperfectly competitive credit markets by supposing that an alternative lender provides an outside option worth a share ϕ of the total surplus created by their lending contract. If $\phi = 1$, then all of the surplus over and above the entrepreneur's outside option accrues to the entrepreneur rather than the lender. This is the competitive benchmark.[13] On the other hand, if ϕ is small, then the lender has a lot of market power.

Timing The timing of production for a type (a, θ) is as follows.

1. Workers choose whether to become an entrepreneur or worker.

---

[11] A similar condition is required in the standard textbook two input profit-maximization problem of the following nature $\max_{k,l} = Ak^\alpha l^\beta - rk - wl$ with $\alpha + \beta < 1$. Namely, a sufficient condition for the second-order condition to hold globally is $\frac{1-\alpha}{\alpha} \frac{1-\beta}{\beta} > 1$.

[12] We can rewrite $\pi(k; \theta, \mathbf{p}) - \delta k > 0$ as $\frac{\pi(k; \theta, \mathbf{p})}{k} > \delta$. As $\pi(k; \theta, \mathbf{p})$ is concave (due to diminishing returns with respect to capital) and $\pi(0; \theta, \mathbf{p}) = 0$:

$$\pi_k(k; \theta, \mathbf{p}) > \frac{\pi(k; \theta, \mathbf{p})}{k}.$$

Therefore at the first-best it must be the case that $\pi_k(k; \theta, \mathbf{p}) > \delta$, which together with the second first-order condition and $g(e; \theta) < 1$ implies:
$$\pi_k(k; \theta, \mathbf{p}) > \gamma > \delta.$$

[13] It could also represent the case where lenders are not-for-profit NGOs or government banks.

2a. If she chooses to become a worker, she inelastically supplies one unit of labour to the labour market.

2b. If she is an entrepreneur, then each lender offers her a contract $(x, r, c, \psi a)$. After deciding whether to accept this contract, she chooses the level of managerial input, e.

3a. With probability $g(e; \theta)$, the firm is successful and then she chooses how much labour to hire, l. Output is realized, wages are paid to managers and workers, and the loan repayment, r, is made.

3b. With probability $1 - g(e; \theta)$, an entrepreneur produces nothing and forfeits collateral, c.

We now work backwards through these decisions to determine the optimal contract. Here, we suppose that the prices, $\mathbf{p}$, are fixed. We then explore the general equilibrium where these are determined.

Labour Hiring With probability $g(e; \theta)$, the firm is successful in which case it decides how many workers to hire to maximize profits, i.e.

$$l^*(k; \theta, \mathbf{p}) = \arg \max_l \left\{ p_y f(k, l; \theta) - p_l l \right\} \tag{4}$$

and define $\pi(k; \theta, \mathbf{p}) := f(k, l^*(k; \theta, \mathbf{p}); \theta) - p_l l^*(k; \theta, \mathbf{p})$ as the conditional profit function given an allocation of capital k. Throughout we make the following assumption, that ensures well-defined interior solutions.

Assumption 2 *The following conditions hold for $g(e; \theta)$ and $\pi(k; \theta, \mathbf{p})$:*

(i) $\pi(k; \theta, \mathbf{p})$ *is strictly concave for all $k \in \mathbb{R}^+$.*

(ii) $g(e; \theta)(\pi(k; \theta, \mathbf{p}) - \delta k)$ *is strictly concave for all $(e, k) \in [0, 1] \times \mathbb{R}^+$.*

(iii) $\lim_{e \to 0} g_e(e; \theta)(\pi(k; \theta, \mathbf{p}) - \delta k) - (1 + g(e)\epsilon(e))p_m > 0$ *for all $k > 0$;*
$\lim_{k \to 0} g(e; \theta)(\pi_k(k; \theta, \mathbf{p}) - \delta) > \gamma - \tau_2 \delta$ *for all $e > 0$.*

These regularity assumptions guarantee that there is a unique global maximum level of managerial input and capital with an interior solution. The last part of the assumption are Inada-like conditions. They are satisfied by the constant elasticity model used in the calibration below.

Choice of Managerial Input We allow lenders to offer credit to entrepreneurs which are tailored to an entrepreneur's characteristics, (a, θ). Since managerial input is costly and unobserved, the level of such input chosen by the entrepreneur has to be incentive compatible.

The expected payoff of an entrepreneur who borrows under contract $\mathbf{t}$ is given by:

$$V(e, \mathbf{t}; a, \theta, \mathbf{p}) = g(e; \theta)(\pi(x + \psi a; \theta, \mathbf{p}) - r) + [1 - g(e; \theta)](\delta(x + \psi a) - c) - p_m e + \gamma(1 - \psi) a. \tag{5}$$

This reflects the fact that, with probability $g(e; \theta)$, the lender is repaid and with probability $(1 - g(e; \theta))$ there is default in which case the lender seizes the entrepreneur's collateral. This is decreasing in the amount of collateral, all else equal. The borrower receives returns on part

of her assets that are not invested in the project and not collected as collateral by the lender in the event of the project failing, namely $\gamma (1 - \psi) a$.

The first-order condition for managerial input is:

$$g_e (e; \theta) [\pi (x + \psi a; \theta, \mathbf{p}) - \delta (x + \psi a) - r + c] = p_m. \tag{6}$$

The level of such input is increasing in collateral c, equity ψa, and the amount borrowed, x. However, it is decreasing in r all else equal, i.e. asking for a higher loan repayment blunts incentives and increases the default rate. Equation (6) is an incentive-compatibility constraint on credit contracts.

Workers who are employed as managers face a risk since the firm may turn out to be unsuccessful. The managerial wage rate must therefore be set such that: $p_m = p_l / g (e, \theta)$ which will vary with e reflecting the fact that riskier firms will have to pay managers a higher premium when hiring managers.

Acceptable Credit Contracts   The limited liability constraint (LLC) with respect to r says that what the lender can take from the borrower is restricted by the net profits of the firm in the event of success plus the expected liquidation value of the borrowers' assets that are not invested in the project:

$$r \leq \pi (x + \psi a) + \tau_1 \gamma (1 - \psi) a. \tag{7}$$

We do not expect this constraint to necessarily bind as the lender has to respect the participation constraint of the borrower. Also, there is the incentive-compatibility constraint, to be formally introduced below, that takes into account the effect of the contractual terms on the borrower's choice of effort and a high value of r will tend to reduce e.

The LLC with respect to c is

$$c \leq \tau_1 \gamma (1 - \psi) a + \tau_2 \delta (x + \psi a). \tag{8}$$

As much as choosing a high value r reduces e, choosing a high value of c tends to increase e. Therefore, this will be the relevant constraint for most of our analysis. We show below that if this constraint does not bind, we will have the first-best. This is intuitive, since with all parties being risk neutral, full residual-claimancy (which in this context means $r = c$) gives the efficient level of e and so would be chosen by the lender if consistent with profit-maximization and feasible given the various constraints.

As well as the level of managerial input being incentive compatible, entrepreneurs must choose to enter lending contracts voluntarily at stage 2, i.e. the contract offered to an entrepreneur of type (a, θ) must generate a payoff which exceeds what is available elsewhere which we denote by u. The yields a participation constraint:

$$V(e, \mathbf{t}; a, \theta, \mathbf{p}) \geq u(a, \theta, \mathbf{p}). \tag{9}$$

In equilibrium, u is determined endogenously and depends on θ, a and $\mathbf{p}$. It can be thought of as a price which endogenously clears the credit market given outside opportunities available to an entrepreneur. In other words, it determines the expected returns from entrepreneurship striking a balance between the demand and supply for different occupations in the economy, which in turn depends on economic fundamentals, such as the distribution of talent and wealth and prices. Below we will determine p_m and p_l endogenously but all individuals take prices as given when making their decisions.

3.3 Credit Contracts in Partial Equilibrium

In this section, we explore access to credit holding fixed who decides to become an entrepreneur and the price vector $\mathbf{p}$. We characterize the form of optimal lending contract $\mathbf{t} = (x, r, c, \psi a)$ by considering two scenarios, i.e. the participation constraint is binding or not.

We begin with a key observation on the properties of such contracts.

Lemma 1 *Under the optimal contract $\psi = 1$.*

The lender will always choose $\psi = 1$, i.e. the borrower's equity participation is at the highest possible level. This is driven by the assumption that in our model outside collateral can be transformed into equity at no cost. The entrepreneur's resources are more valuable as equity – which allows to reduce the loan amount – than as collateral, given the inefficiencies associated with lending contracts. As a result there is only inside collateral and the contracting friction τ_1 does not matter for the allocation. The limited liability constraint given $\psi = 1$ is

$$c \leq \tau_2 \delta (x + a).$$

Note that the limited liability constraint will always be binding in the second best case, i.e. as long as $c < r$. The reason is that increasing the collateral value dominates increasing the repayment burden in case of success: both transfer resources to the lender, but the former has a positive incentive effect, while the latter has a negative incentive effect.

When the Participation Constraint is not Binding Suppose the participation constraint is not binding, then using the incentive compatibility constraint for the borrower and the fact that $k = x + a$, we can rewrite from (3) the optimal contracting problem as[14]:

$$\max_{e,k} g(e; \theta) \left(\pi(k; \theta, \mathbf{p}) - \frac{p_m}{g_e(e; \theta)} - \delta k \right) + \tau_2 \delta k - \gamma(k - a). \tag{10}$$

The first-order necessary conditions characterizing an interior optimum $(e_0(\theta, \mathbf{p}), k_0(\theta, \mathbf{p}))$ are

$$g_e(e_0) \left[\pi(k_0; \theta, \mathbf{p}) - \delta k_0 \right] = \left[1 + g(e_0; \theta) \epsilon(e_0; \theta) \right] p_m \tag{11}$$

$$g(e_0) \left[\pi_k(k_0; \theta, \mathbf{p}) - \delta \right] = \gamma - \tau_2 \delta \tag{12}$$

By Assumption 2(iii), the unique global maximum (e_0, k_0) is an interior solution. Note that in this case, the optimal managerial input, $e_0(\theta, \mathbf{p})$, and capital level, $k_0(\theta, \mathbf{p})$, are independent of u. The intuition behind equation (11) is that, it is "as if" the cost of managerial input is increased by the term $g(e_0; \theta) \epsilon(e_0; \theta)$ which represents the marginal "agency cost" due to moral hazard.

Equation (12) determines capital allocation. Two forces lead to an allocation of capital that is different from first best. First, the firm's capital is depressed as default risk increases. However, conditional on default probabilities, this re-allocation of capital is efficient in the sense that expected marginal returns are equalized. Second, firm capital will be depressed when $\tau_2 < 1$. This distortion is caused by the lenders' anticipation that they will not be able to recover the full collateral in case of default, effectively raising the costs of funds. Equation (12) emphasizes the

---

[14]We have shown in the appendix that in the second best case, the limited liability constraint is binding, i.e. $c = \tau_2 \delta k$., in a way similar to Besley et al. (2012).

role that equilibrium default has on the capital available to a firm. Unlike most of the existing literature, (e.g., Buera et al 2011, 2015), the credit market friction affecting capital allocation is determined in equilibrium as a function of the equilibrium price vector and outside option in addition to borrower characteristics a and θ. This will also be a feature of the calibration of the model below and we will explore heterogeneity in default rates in this setting.

Any credit contract with $(e_0(\theta, \mathbf{p}), k_0(\theta, \mathbf{p}))$ yields the same surplus, and we will show that it is the optimal contract when the borrower has an outside option below some threshold $\underline{u}(\theta, \mathbf{p})$.

When the Participation Constraint is Binding Using the binding participation constraint and the incentive compatibility constraint, we can characterize the implicit equation characterizing the borrower's optimal effort level as

$$p_m \left[\frac{g(e; \theta)}{g_e(e; \theta)} - e \right] + (1 - \tau_2) \delta k = u. \tag{13}$$

This gives the optimal effort level as a function of k, $e^* = \nu(u, k; \theta, \mathbf{p})$, and by partial differentiation, we know $\nu(u, k; \theta, \mathbf{p})$ is increasing in u and decreasing in k.

Now using the binding participation constraint, the optimal contracting problem becomes

$$\max_k g\left(\nu(u, k; \theta, \mathbf{p})\right) \left(\pi(k; \theta, \mathbf{p}) - \delta k\right) + \delta k - p_m \nu(u, k; \theta, \mathbf{p}) - \gamma(k - a) - u$$

The first-order necessary condition characterizing the optimal capital level k^* is given by

$$g\left(\nu(u, k^*; \theta, \mathbf{p})\right) \left(\pi_k(k^*; \theta, \mathbf{p}) - \delta\right) + \nu_k(u, k^*; \theta, \mathbf{p}) \left[g_e\left(\nu(u, k^*; \theta, \mathbf{p})\right) \left(\pi(k^*; \theta, \mathbf{p}) - \delta k^*\right) - p_m\right] = \gamma - \delta \tag{14}$$

This equation captures the direct effect of equilibrium default on capital allocation, i.e. the same term $g(e; \theta)(\pi_k(k; \theta; \mathbf{p}) - \delta)$ also appeared in equation (12). In addition, higher capital will increase default and hence decrease surplus, which is the term in square brackets.[15] This is because the increase in capital corresponds to lump-sum transfer to the borrower: in the case of default δk can be liquidated, but this can only partly be recovered by the lender as long as $\tau_2 < 1$. The lender will want to increase the repayment burden over and above what he would do to extract additional surplus generated by additional capital, and this depresses managerial input. Note that this distortion is amplified the lower is τ_2, as can be seen from equation (13): for $\tau_2 < 1$ any increase in capital will only partially increase the collateral value of the project, which requires an even higher increase in the repayment burden to be beneficial to the lender; that however is detrimental to the provision of managerial input.

Further note that the optimal capital level can be written as an increasing function of outside option, $\varsigma(u; \theta, \mathbf{p})$. The optimal managerial input level is defined as $\xi(u; \theta, \mathbf{p}) := \nu(u, \varsigma(u; \theta, \mathbf{p}); \theta, \mathbf{p})$. We show in the appendix that both the capital level and the optimal managerial input are interior solutions, and both are increasing function of u. Any improvement in an entrepreneur's outside option will require the lender to impose a lower repayment burden, which increases managerial inputs, reduces default, and consequently increases the level of capital in the firm.

Note that when u is large enough, i.e. $u \geq \overline{u} = p_m \left(\frac{g(e_{FB}; \theta)}{g_e(e_{FB}; \theta)} - e_{FB} \right) + (1 - \tau_2) \delta k_{FB}$, the first best outcome is achieved, where managerial input level is chosen to set the marginal benefit

---

[15]It can be shown that the term in square brackets is positive. Further note that $\nu_k(u, k^*; \theta, \mathbf{p}) < 0$.

equal to the marginal cost when the entrepreneur is a full residual claimant. At the other extreme, for low u, the participation constraint will not binding. The value of u in determining the optimal contract is summarized in our next result:

Proposition 1 *There exists $[\underline{u}(\theta, \mathbf{p}), \overline{u}(\theta, \mathbf{p})]$ such that optimal lending contracts yield managerial input $\hat{e}$ and total lending $\hat{x}$, as follows:*

$$
\hat{e}(u; \theta, \mathbf{p}) = \begin{cases} e_0(\theta, \mathbf{p}) & \text{for } u < \underline{u}(\theta, \mathbf{p}) \\ \xi(u; \theta, \mathbf{p}) & \text{for } \underline{u}(\theta, \mathbf{p}) \leq u < \overline{u}(\theta, \mathbf{p}) \\ e_{FB}(\theta, \mathbf{p}) & \text{for } u \geq \overline{u}(\theta, \mathbf{p}) \end{cases}
$$

where $e_0(\theta, \mathbf{p})$ is a constant, defined in the case where the participation constraint is not binding; $e_{FB}(\theta, \mathbf{p})$ is a constant equal to first best managerial input; $\lim_{u \to \underline{u}(\theta, \mathbf{p})} \xi(u; \theta, \mathbf{p}) = e_0(\theta, \mathbf{p})$ and $\lim_{u \to \overline{u}(\theta, \mathbf{p})} \xi(u; \theta, \mathbf{p}) = e_{FB}(\theta, \mathbf{p})$.

$$
\hat{x}(u; a, \theta, \mathbf{p}) = \begin{cases} k_0(\theta, \mathbf{p}) - a & \text{for } u < \underline{u}(\theta, \mathbf{p}) \\ \varsigma(u; \theta, \mathbf{p}) - a & \text{for } \underline{u}(\theta, \mathbf{p}) \leq u < \overline{u}(\theta, \mathbf{p}) \\ k_{FB}(\theta, \mathbf{p}) - a & \text{for } u \geq \overline{u}(\theta, \mathbf{p}) \end{cases}
$$

where $k_0(\theta, \mathbf{p})$ is a constant, defined in the case where the participation constraint is not binding; $k_{FB}(\theta, \mathbf{p})$ is a constant equal to first best managerial input; $\lim_{u \to \underline{u}(\theta, \mathbf{p})} \varsigma(u; \theta, \mathbf{p}) = k_0(\theta, \mathbf{p})$ and $\lim_{u \to \overline{u}(\theta, \mathbf{p})} \varsigma(u; \theta, \mathbf{p}) = k_{FB}(\theta, \mathbf{p})$.

The gross repayment r is pinned down by the binding incentive compatibility constraint and limited liability constraint as

$$
r(u; a, \theta, \mathbf{p}) = \pi(\hat{x}(u; a, \theta, \mathbf{p}) + a; \theta, \mathbf{p}) - (1 - \tau_2)\delta(\hat{x}(u; a, \theta, \mathbf{p}) + a) - \frac{p_m}{g_e(\hat{e}(u; \theta, \mathbf{p}); \theta)}. \quad (15)
$$

Notice that while $\hat{x}(u; a, \theta, \mathbf{p})$ does depend on wealth a, the optimal capital level $\hat{k} := \hat{x}(u; a, \theta, \mathbf{p}) + a$ does not. We can then define the indirect total surplus in the lending relationship as

$$
\hat{S}(u; \theta, \mathbf{p}) := S(\hat{e}(u; \theta, \mathbf{p}), \hat{x}(u; a, \theta, \mathbf{p}) + a; \theta, \mathbf{p}) \quad (16)
$$

Observe that surplus does not depend on wealth a directly, other than through any effect it might have on u. Further, for the first best case and the case where participation constraint is not binding, the optimal managerial input and capital level are independent from u, so u will not affect total surplus, i.e. $S_u = 0$, where S_u denotes the partial derivative of the surplus function with respect to u. The following result gives a characterization of the ranges in which u can fall in terms of the surplus function:

Corollary 1 *The indirect surplus function, $\hat{S}(u; \theta, \mathbf{p})$, is increasing in u for all $u \in (\underline{u}(\theta, \mathbf{p}), \overline{u}(\theta, \mathbf{p}))$. For $u \geq \overline{u}(\theta, \mathbf{p})$ or $u \leq \underline{u}(\theta, \mathbf{p})$, $\hat{S}(u; \theta, \mathbf{p})$ is constant.*

Credit contracts will implement first best level of managerial input as long as entrepreneurs can provide sufficient collateral, i.e. has high a, or a high outside option. The first of these is standard feature of existing models of *ex post* enforcement constraints. What the general equilibrium contracting model emphasizes is that whether the first-best is attainable also depends on an endogenously determined outside option which affects the equilibrium default rate. In the intermediate u range, greater collateral allows for more efficient lending since it

relaxes (6). The lender then offers a higher x, which amplifies the effect of collateral on the incentive compatibility constraint. Similarly, a higher outside option increases lending efficiency. The lender has to transfer a greater share of surplus to the entrepreneur, and this is optimally implemented by reducing r and increasing x, which in turn increase managerial input. However, for $u \leq \underline{u}(\theta, \mathbf{p})$ the lender will always implement $e_0(\theta, \mathbf{p})$. In this range – due to the concavity of $g(e; \theta)$ – a reduction in r increases surplus by more than it transfers surplus to the entrepreneur. Therefore it is in the interest of the lender to offer a contract which leaves the entrepreneur with an expected income greater than the outside option. It is optimal to transfer surplus by decreasing r. In this region, the lender reacts to an increased c by increasing r by the same amount, and leaving both e and x unchanged. Surplus stays unchanged, but is transferred from the borrower to the lender.

The Lender's Participation Constraint   Whether a lender wishes to lend to an entrepreneur of type (a, θ) depends upon whether they can make a profit by doing so.   Hence for an entrepreneur of type (a, θ) to be offered any credit requires that

$$\hat{\Pi}(u; a, \theta, \mathbf{p}) := \Pi(\hat{e}(u; \theta, \mathbf{p}), \mathbf{t}(u; a, \theta, \mathbf{p}); \theta) \geq 0.$$

Determining the Entrepreneur's Outside Option   The final part of the partial equilibrium analysis is to determine the entrepreneur's outside option endogenously.   This will be the maximum of three things: (i) what she can obtain by borrowing from another lender, (ii) self-financing the project with the (limited) wealth owned and (iii) working for a wage. We now explore this in detail.

Let $\hat{u}(\phi; a, \theta, \mathbf{p})$ be defined by:

$$\phi \cdot \hat{S}(\hat{u}(\phi; \theta, \mathbf{p}); \theta, \mathbf{p}) = \hat{u}(\phi; \theta, \mathbf{p}). \tag{17}$$

This implicitly defines the equilibrium payoff of an entrepreneur if the only outside option is to receives a share ϕ of the surplus in a lending relationship. Note that this is not the payoff from borrowing since the efficiency utility in Proposition 1 bounds the borrower's payoff from below when the outside option is low, and in particular when ϕ is low.[16]

Now consider the payoff where the agent chooses to self-finance, i.e. use only his own wealth. This is given by

$$V^{self}(a, \theta, \mathbf{p}) = \max_{(e,k)} \left[g(e; \theta) \left(\pi(k; \theta, \mathbf{p}) - \delta \right) - p_m e + \gamma(a - k) : k \leq a \right]. \tag{18}$$

Let $\{ e^{self}(a, \theta, \mathbf{p}), k^{self}(a, \theta, \mathbf{p}) \}$ denote the solutions to the maximization problem (18). Lastly the entrepreneur could choose to become a wage labourer. The entrepreneurs outside option will therefore be given by

$$u(a, \theta, \mathbf{p}) = \max\{ V^{self}(a, \theta, \mathbf{p}), \hat{u}(\phi; \theta, \mathbf{p}), p_l + \gamma a \}.$$

Comparative Statics   We now have the following result for the payoff of entrepreneurs:

---

[16]Note that even with $\phi = 0$, the lender does not necessarily receive u since, as we observed Proposition 1, the entrepreneur's participation constraint might not be binding.

Proposition 2 *The entrepreneur's expected profit increases with more competition (ϕ) and greater wealth (a). Without further assumptions, the effect of productivity (θ) on entrepreneur's expected profit is indeterminate.*

Thus entrepreneurs benefit from increased competition since they get a larger share of the surplus in the credit market. They also do better when they have more collateral to post. Increasing productivity has competing effects which explains the ambiguous effect on total surplus. On the one hand, profits are higher as firms are more productive. However, the effect on the repayment probability is ambiguous since the cost of managerial input is larger in larger firms.

3.4 General Equilibrium

So far, we have taken the price vector $\mathbf{p}$ and the occupational structure as given. Our general equilibrium analysis determines these endogenously.

Financial Market Access We assume that a fraction $z(a, \theta) \in [0, 1]$ of agents of type (a, θ) has access to financial markets.[17] Denote with $\chi \in \{0, 1\}$ whether any given individual has access to credit markets. Let $h(a, \theta)$ denote the joint density associated with the distribution of (a, θ). Total financial inclusion in the economy is defined by

$$\bar{\chi} := \int \int z(a, \theta) h(a, \theta) \, da \, d\theta,$$

i.e. as the proportion of agents who have market access. If they have access then they can access credit markets as described in the previous section.

Occupational Choice Let $\sigma \in \{0, 1\}$ denote whether an agent becomes an entrepreneur, with $\sigma = 1$ indicating entrepreneurship and $\sigma = 0$ indicating becoming a worker. An agent will choose entrepreneurship when the expected payoff from being an entrepreneur exceeds that from being a wage labourer. Formally,

$$\sigma(\chi, a, \theta, \mathbf{p}) = \begin{cases} 1 & \text{if } \chi = 1 \text{ and } \hat{\Pi}(u(a, \theta, \mathbf{p}); a, \theta, \mathbf{p}) \geq 0, \text{ or, if } V^{self}(a, \theta, \mathbf{p}) \geq p_l + \gamma a. \\ 0 & \text{otherwise.} \end{cases}$$

The borrower will always choose to become an entrepreneur if the autarky payoff is bigger than the wage. If she has access to credit markets, she will also become an entrepreneur if the lender can offer a profitable credit contract (satisfying the borrower's outside option and incentive constraint). Clearly this depends on the individuals type (a, θ). Moreover, since the payoff from entrepreneurship is increasing in a and θ, if a type (a, θ) becomes an entrepreneur then so do all individuals with higher wealth and productivity. Hence, there will be critical values of wealth and productivity that define the entrepreneurial class. How dense this is depends on the joint distribution of wealth and productivity.

---

[17]We assume that aurtarky is the only alternative to credit market access. An interesting extension in future work would be to allow an informal sector which could be characterized by a higher cost of funds, γ and would be another potential outside option for the borrower.

Equilibrium Wages To determine equilibrium wages, we need to solve for aggregate labour supply and demand in the economy. This means aggregating over the distribution of wealth and productivity. Aggregate labour supply is determined by the fraction of individuals who choose not to become entrepreneurs, i.e.

$$L^S(\mathbf{p}) = \int \int [z(a,\theta)\{1 - \sigma(1,\theta,a,\mathbf{p})\} + (1 - z(a,\theta))\{1 - \sigma(0,\theta,a,\mathbf{p})\}]\, h(a,\theta)\, da\, d\theta. \quad (19)$$

Denote managerial labour demand, conditional on becoming entrepreneur, as

$$\hat{e}(\chi,a,\theta,\mathbf{p}) := \chi(\hat{e}(u(a,\theta,\mathbf{p}); a,\theta,\mathbf{p}) + (1 - \chi)e^{self}(a,\theta,\mathbf{p}),$$

and firm capital, conditional on becoming entrepreneur, by

$$\hat{k}(\chi,a,\theta,\mathbf{p}) := \chi(\hat{k}(u(a,\theta,\mathbf{p}); a,\theta,\mathbf{p}) + (1 - \chi)k^{self}(a,\theta,\mathbf{p}).$$

To solve for aggregate labour demand we need to take into account the fraction of firms that are operational given the equilibrium default probability which we denote by

$$\hat{g}(\chi,a,\theta,\mathbf{p}) = g(\hat{e}(\chi,a,\theta,\mathbf{p});\theta).$$

Note that this also depends on $\mathbf{p}$ through its effect on profits and the cost of managerial input. Labour demand also depends on the amount of labour hired by each firm, conditional on producing. We will denote this by

$$\hat{l}(\chi,a,\theta,\mathbf{p}) = l^*(\hat{k}(\chi,a,\theta,\mathbf{p});\theta,\mathbf{p})$$

using (4).

Aggregate labour demand is then given by

$$\begin{aligned}
L^D(\mathbf{p}) &= \int \int z(a,\theta)\left[\sigma(1,a,\theta,\mathbf{p}) \cdot \left(\hat{l}(1,a,\theta,p) \cdot \hat{g}(1,a,\theta,\mathbf{p}) + \hat{e}(1,a,\theta,\mathbf{p})\right)\right] h(a,\theta)\, da\, d\theta \\
&+ \int \int (1 - z(a,\theta))\left[\sigma(0,a,\theta,\mathbf{p}) \cdot \left(\hat{l}(0,a,\theta,p) \cdot \hat{g}(0,a,\theta,\mathbf{p}) + \hat{e}(0,a,\theta,\mathbf{p})\right)\right] h(a,\theta)\, da\, d\theta. \quad (20)
\end{aligned}$$

This is the sum over the labour demand functions of individuals who choose to become entrepreneurs at prevailing prices $\mathbf{p}$, characterized by (a,θ,χ), .

The equilibrium prices $\hat{\mathbf{p}}$ now equates supply and demand, i.e. solves

$$L^S(\hat{\mathbf{p}}) = L^D(\hat{\mathbf{p}}).$$

This depends implicitly on all dimensions of choice: occupational choice, credit contracts which determines use of capital and labour demand. It also depends on the extent of financial access since this will affect who becomes an entrepreneur and the amount of labour demand among those who do, depending on whether they can access financial markets.

3.5 Two Benchmarks

Before proceeding to study the calibration of the model, it is worth considering two special cases that will serve as useful benchmarks in what follows: autarky and the first best.

Autarky We define autarky purely in terms of credit markets, i.e. to describe a situation where there is only trade in labour and goods markets, but not in capital. Formally, this is a case where $z(a, \theta) = 0$ for all (a, θ). In this cases, the only way in which individuals can access credit is via their own wealth. The choice of managerial input and capital are given by (18). In autarky there can be wide dispersion in the marginal product of capital across entrepreneurs: an entrepreneurs' firm's capital is constrained by his personal wealth. Associated with autarky will be a wage rate $\mathbf{p}^{aut}$ which clears the labour market given the occupational choice decisions.

By misallocating capital, autarky also results in lower labour demand. This in turn depresses wages. This means that wages will tend to be lower so autarky can actually encourage people to become entrepreneurs compared to a situation where capital markets are functioning well.

The First-Best We now consider what would happen with perfect capital markets. This has two dimensions. First, there is complete access to financial markets, $z(a, \theta) = 1$ for all (a, θ), and second, there is no moral hazard problem. In effect, the latter implies that a lender can specify a level of managerial input as part of the lending contract. Efficient credit contracts are characterized by (1) and (2). Note that the first best does have a level of default associated with it. However, these decisions and payoffs are independent of a, i.e. the entrepreneur's level of wealth is irrelevant.

Occupational choice is given by

$$\sigma_{FB}(\theta, \mathbf{p}) = \begin{cases} 1 & \text{if } S(e_{FB}, k_{FB}; \theta, \mathbf{p}) \geq p_l \\ 0 & \text{otherwise} \end{cases}$$

which is also independent of a. Associated with first-best will be a price vector $\mathbf{p}_{FB}$ which clears the labour market given the occupational choice, capital allocation and labour demand decisions. The wage rate will be endogenous and set to clear the labour market.

4   From Theory to Data

We use the model to produce a range of calibrated counterfactuals to explore the model's predictions quantitatively. The framework allows us to think about two main things. First, we can think about the effect of credit market frictions on optimal credit contracts. We can explore the effect of two specific frictions as represented by ϕ and τ_2. Second, we can look at impact of changing market access as represented by $z(a, \theta)$.

Changing market frictions affects labour demand for a given wage in (20) through three channels. First, it increases access to capital and this increases labour demand since capital and labour are complements. Second, it reduces the default probability by increasing managerial input. Third, it lowers the threshold productivity and wealth levels at which agents choose to become entrepreneurs. Increasing $z(a, \theta)$ has a direct effect on labour demand since some entrepreneurs now get access to more capital.

General equilibrium effects are largely driven by shifts in labour demand and occupation choice which affect the wage which, in turn, feeds back on to the participation constraint of entrepreneurs and hence to the terms of credit contracts. Wages also affect the amount of managerial labour applied by changing profitability and the amount of capital used.

The model is able to give a clear sense of the different "moving parts" that affect credit market frictions in a general equilibrium model with endogenous occupational choice. Our next step is to put the model to work by exploring different aspects of its quantitative predictions. For this, we will need to give a specific parametrization and simulate the model's predictions which will give insights in three main areas.

We next describe the specific functional forms that use and then discuss how various key parameters are calibrated.

4.1 Parametrization

The production function, $f(k, l; \theta)$ is Cobb-Douglas with diminishing returns:

$$f(k, l; \theta) = \theta^{1-\eta-\alpha} \left(l^{1-\beta} k^{\beta} \right)^{\eta}, \tag{21}$$

where θ is the firm specific productivity parameter and $\alpha, \beta,$ and η, all of them belonging to the interval $(0, 1)$, are parameters governing the shape of the production function. Thus the model is essentially a classic Lucas-style "span of control" model η representing the extent of diminishing returns and pure profits can be thought of as payment to an untraded factor such as technology or ability.

Using this, a firm's labour demand, conditional on k, is given by:

$$l^{*}(k; \theta, \mathbf{p}) = \left[\eta(1-\beta) \frac{p_y}{p_l} \theta^{1-\eta-\alpha} k^{\eta\beta} \right]^{\frac{1}{1-\eta(1-\beta)}} \tag{22}$$

and the conditional profit function is

$$\pi(k; \theta, \mathbf{p}) = (1 - \eta(1-\beta)) \left[p_y \left(\frac{\eta(1-\beta)}{p_l} \right)^{\eta(1-\beta)} \theta^{1-\eta-\alpha} k^{\eta\beta} \right]^{\frac{1}{1-\eta(1-\beta)}} \tag{23}$$

The marginal product of capital is therefore given by:

$$\pi_k(k; \theta, \mathbf{p}) = \eta\beta \left[p_y \left(\frac{\eta(1-\beta)}{p_l} \right)^{\eta(1-\beta)} \theta^{1-\eta-\alpha} k^{\eta-1} \right]^{\frac{1}{1-\eta(1-\beta)}} \tag{24}$$

For the success technology, we use a constant-elasticity functional form where:[18]

$$g(e; \theta) = \lambda \left[e/\theta^{\mu} \right]^{\alpha} \text{ with } \mu \geq 0.$$

The parameter μ governs the dependence of the cost of managerial input on θ, i.e. the link between this and firm size. If $\mu = 0$, then the cost of securing a given level of default does not depend on firm size whereas $\mu > 0$ means that achieving the same default in a large firm requires more managerial input. The parameter α in the technology above governs the elasticity of the success probability with respect to managerial input.[19] Together with the assumption

---

[18] An alternative isomorphic specification in terms of effort choice would assume that the same level of effort is more costly in large firms: $g(e; \theta) = \lambda e^{\alpha}$ and cost of effort $p_e \theta^{\mu} e$.

[19] The parameter α will be chosen such that first best default probabilities $g(e^*(\theta, \mathbf{p}); \theta)$ match their empirical counter-part, which is particularly high for large firms, i.e. at the highest level of θ. For lower levels of θ and in second-best default probabilities will be higher, i.e. success probabilities will be lower. Therefore no additional assumption is required to guarantee that $g(e; \theta) \in [0, 1]$.

in (21) this functional form implies that output has constant returns to scale in managerial input (e), capital (k), labour (l), and entrepreneurial talent (θ). Finally, the parameter λ captures the general productivity of managerial input in achieving project success. In the next section, we will show how to use data on the firm size distribution and heterogeneous default probabilities by firm size to calibrate (μ, α, λ). Each agent who works as an employee is indifferent between being a worker (providing input l) and managerial labour; they are paid at rate p_l, or - alternatively - at a risk-adjusted wage rate p_m ($= p_l / g(e; \theta)$) in case of success.

4.2 Calibration

Without loss of generality, we will take the output price to be the same across countries and choose the unit of measurement such that $p_y = 1$. Since we will think of the price of capital goods (but not necessarily the rental rate) to be equal across countries, we measure capital, k in value terms. Further we will assume $p_m / g(e; \theta) = p_l$.

Model Parameters   We calibrate a subset of the model parameters using evidence from existing studies. First, we assume that β, which in first best measures the share of output paid to capital relative to labour[20], is $1/3$ in line with standard calibrations used in the macroeconomic literature. Second, we set η to $3/4$, following the assumption of Bloom (2009) in a related context. Third, we assume $\lambda = 1.05$ for the calibration while in the simulations we will set $\lambda = 1.0$. This assumes that entrepreneurial productivity is 5% higher in the US than in the simulated economy. This is broadly consistent with evidence linking entrepreneurial quality with firm survival. For example, Bloom and Van Reenen (2010) find the lowest level of management performance in China, India, Greece and Brazil, and the highest level of management performance in the United States. Consistent with our model, they find that in the cross section of firms their measure of management performance is associated with significantly higher firm survival. Quantitatively their results suggests that moving from the average management quality in China or India to the average management quality in the US will increase default rates by about 0.35 percentage points over a horizon of about 5 years. Fourth, we take the marginal cost of capital γ to be 1.02, which roughly corresponds to long run real interest rates in the US since the 1980's (Yi and Zhang, 2016) with an allowance for capital depreciation.[21] Fifth, δ is the recovery rate in case of default. We back this out from data on the charge-off rate (0.0084) and delinquency rate (0.0279) for corporate loans since 1985 in the US. Those imply $\delta = 0.711$, or a loss of 28.9% of the loan value in case of corporate defaults.[22]

   We normalize, without loss of generality, the US wage to be one, and therefore any wage or income level in the distorted model is measured relative to the US wage.

Entrepreneurship and the Distribution of Productivity   The remaining parameters are α and μ and the distribution of θ. We pin down those parameters jointly by calibrating the model to US data, assuming that this is an example of "perfectly" functioning credit markets. While this assumption is somewhat extreme, it may still serve as a reasonable approximation of the

---

[20]Note that this only holds when defining the labor income share as payments to l, not e.

[21]The rates paid to depositors in developing countries are of little guidance to calibrate γ if depositors are not the marginal source of funding, or there are transaction costs in financial intermediation.

[22]Data from Board of Governors of the Federal Reserve (2016, https://www.federalreserve.gov/releases/chargeoff/).

difference between US credit markets and credit markets in developing countries which is our main focus of attention. What makes this assumption convenient is that all of the model's predictions are independent of the wealth distribution, allowing to calibrate the unknown parameters without knowledge of the wealth distribution. We can then specify any hypothetical marginal wealth distribution and its correlation with the productivity distribution when we simulate second best outcomes.

The parameters α and μ and the distribution of θ jointly determine the pattern of corporate default rates across firm sizes and the firm size distribution and therefore need to be chosen *jointly* to match those moments in the US data. However, to clarify which variation in the data pins down which parameters, it is instructive to consider how the pattern of corporate default rates across firm sizes depends on α and μ, conditional on the θ distribution; and how the firm size distribution is linked to the θ distribution conditional on α and μ.

First, we can solve or the first best level of managerial labour e_{FB} in closed form as:

$$e_{FB}(\theta, \mathbf{p}) = \left[p_y \theta^{1-\eta-\alpha} \left(\frac{\eta(1-\beta)}{p_l} \right)^{\eta(1-\beta)} \left(\frac{\eta\beta}{\gamma} \right)^{\eta\beta} \left(\frac{\lambda\alpha(1-\eta(1-\beta))}{p_l \theta^\mu} \right)^{1-\eta} \right]^{\frac{1}{(1-\alpha)(1-\eta)-\alpha\eta\beta}} \tag{25}$$

Generally a tuple (μ, α) implies that first best default probabilities increase or decrease with first best firm size: the parameter α govern the general level of default, and the parameters μ governs the rate at which default probabilities increase/decrease with firm size. We choose α and μ such that the (loan value weighted) average default probability amongst firms with less than 250 employees and firms with more than 250 employees matches the average default probabilities of 0.105 and 0.062 for those respective firm size categories reported in an early version of Besley, Roland, van Reenen (2020).

Second, the marginal distribution of θ can be backed out from data on the distribution of firms sizes, conditional on α and μ. Plugging $k_{FB}(\theta, \mathbf{p})$ into (22) we can write equilibrium labour demand, $l_{FB}(\theta, \mathbf{p}) := l^*(k_{FB}; \theta, \mathbf{p})$ as a monotonically increasing function of θ. Empirically the distribution of firm sizes measured in terms of the size of the labour force. This corresponds to $e_{FB}(\theta, \mathbf{p}) + l_{FB}(\theta, \mathbf{p})$ in the model, assuming that only successful firms are recorded in the data. Both $e_{FB}(\theta, \mathbf{p})$, given the above calibration, and $l_{FB}(\theta, \mathbf{p})$ are monotonically increasing in θ. Empirically the firm size distribution is known to be well approximated by a Pareto distribution, with shape parameter $\sigma_l = 1.059$ (Axtell, 2001). The monotonic relationship described above allows, for any size of the labour force observed in the data, to back out θ.[23] Further we can calculate the associated probability of success and hence back out the density of that θ in the population given the empirical frequency of that firm size.

However, this procedure only works for any θ such that $\sigma(a, \theta, \mathbf{p}) = 1$, i.e. values of θ for which individuals choose to operate the project. To pin down the full θ distribution we assume that also the distribution of successful firms that would be observed if all individuals decided to operate their project follows a Pareto distribution with shape parameter $\sigma_l = 1.059$. Note that this assumption is consistent with the data, since a lower-truncated Pareto distribution is again a Pareto distribution with the same shape parameter. We then find the scale parameter of that firm size distribution, and more importantly the associated full θ distribution, such that labour markets clear at the observed US wage, i.e. solve $L^S(\mathbf{p}) = L^D(\mathbf{p})$ at $p_l = 1$.

---

[23]This implies that in the first best scenario, the largest firm in our simulations is as large as the largest US firm in our data. In second best simulations larger firms might emerge to the extend that high productivity entrepreneurs have access to capital and wages are depressed.

The Distribution of Wealth We can specify the marginal asset distribution to follow any observed or hypothetical wealth distribution. For our baseline simulations we choose the marginal distribution of assets to approximate the wealth distribution in India. We obtained data on the Indian wealth distribution from the Global Wealth Report 2015 (Credit Suisse, 2015).

This report provides information on the Gini coefficient of the Indian wealth distribution, mean wealth, median wealth and the fraction of the population in four wealth classes: 0-10fk, 10k-100k, 100k-1m and over 10m USD. The median wealth in India is 1.75% of median wealth in the US, and the mean wealth is 1.24% of mean wealth in the US.

We assume the Indian wealth distribution to be of the Pareto family, which has been shown to be a reasonable approximation in a number of countries. This reduces the calibration to choosing a shape and scale parameter of that distribution. Moreover, given the Pareto assumption, the shape parameter has a known monotonic relation to the Gini coefficient. We use this relation together with the aforementioned data on the empirical Gini coefficient to back out the shape parameter. Specifically, the scale parameter is chosen to minimize the sum of squared differences between the empirical probability mass and the probability mass of the calibrated Pareto distribution in each of the four wealth categories, where the summation is across wealth categories.

Lastly, we need to specify the joint distribution $h(a, \theta)$ of assets and productivities.[24] This is difficult to back out non-parametrically from data. In a world with first best credit contracts, knowledge of individual wealth levels, occupational status, and the size of the labour force of firms held by entrepreneurs, would be sufficient to back out the joint distribution of a and θ for the subset of individuals with a θ high enough to become entrepreneurs. However, for all individuals with a value of θ that does not lead to them becoming entrepreneurs, θ is fundamentally unobserved. In our simulations we therefore work with several hypothetical joint distributions.

To this end, we can specify a pattern of dependency between a and θ using the statistical concept of copulas.[25] According to Sklar's theorem (Sklar, 1959), the multivariate density function $h(a, \theta)$ can be rewritten as $h(a, \theta) = h_a(a) \cdot c(H_a(a), H_\theta(\theta)) \cdot h_\theta(\theta)$, where $H_a(\cdot)$ and $H_\theta(\cdot)$ are the cumulative density functions of the marginal distribution of a and θ, respectively, $h_a(\cdot)$ and $h_\theta(\cdot)$ are the corresponding probability density functions, and $c : [0, 1]^2 \to R^+$ is the density function of the copula. We assume that the dependency between a and θ is characterized by a Normal copula. This implies that the only free parameters that have to be specified is the covariance which we choose such that the induced correlation between a and θ matches one of a range of "target values" of the correlation: $\rho \in \{0.0, 0.3\}$. As we increase ρ we are postulating a stronger link between productivity and wealth. Using this approach, we can simulate our model given each value of ρ to trace out the implications of different degrees of correlation between a and θ for credit market outcomes.

---

[24]Here we take the joint distribution of productivity and wealth as a primitive. This contrasts with the approach taken in Buera, et al (2011, 2017) where the distribution of productivity is the only primitive, $h(a, \theta)$ then being determined endogenously through the agents' saving behavior. In a model with default, we would not expect the distribution of wealth to be pinned down only by θ since it would depend on the history of default which would wipe out an entrepreneur's wealth in our framework. Introducing savings into our framework is an important future extension. More generally, the framework that we are proposing could handle shocks to the value of wealth due, for example, to asset price fluctuations which hit agents heterogeneously.

[25]See Nelsen (1999) and Trivedi and Zimmer (2007) for accessible introductions.

4.3 Computation

In order to compute the model, we approximate the continuous distribution of a and θ by a distribution with 1000 and 5000 discrete values, respectively, both in the calibration and the subsequent simulations. These discrete values represent equally spaced centiles of the continuous distribution.

When calibrating the model we solve jointly for the distribution of productivities, α, μ using an iterative process as follows. We start from an initial trial value of the parameters affecting default risk and the cost of managerial input, (μ, α), and then find the distribution of productivity, θ, to match the empirical firm size distribution and ensure that the labour markets clear at a wage (p_l) of one as described above. Conditional on this distribution of θ we then update the value of (μ, α) to generate the empirical average default probabilities for small and large firms which are active in the equilibrium. We then iterate this process until the values of both α and μ converge in the sense that their values change each by less than 0.1 percentage points relative to the previous iterations.

The core problem of the simulations is to find the equilibrium wage at each level of ρ, τ_2 and ϕ. We implement this computationally using the bisection method. A wage is accepted as a solution once labour demand relative to labour supply deviates by less than 0.001 from 1. Given any ρ, τ_2 and ϕ and wage, the simulations involve computing the credit contracts for each of the 1000×5000 tuples for (a, θ). In order to speed up the computation, we make use of the result that if a potential entrepreneur decides to become a worker at (a, θ), all individuals with the same productivity and lower wealth will also choose to become workers.

5 Results

For the results that follow, we will consider an economy where the productivity distribution is based on the US and the wealth distribution on India as detailed in the previous section. The benchmark that we study has no correlation between wealth and productivity ($\rho = 0$). For the core results presented here, we set $\lambda = 1$ so that our US benchmark is 5% more productive than the economy that we are studying translating managerial input into repayment success. We then set $\tau_2 = 1$ so that property rights to wealth are perfect, i.e. all wealth can be used as collateral. All of these assumptions will be maintained in what follows unless we explicitly state otherwise. Capital can be acquired by lenders at borrowing rate of 10% so that $\gamma = 1.02$.[26] In the first best, the marginal product of capital will be equal to this.
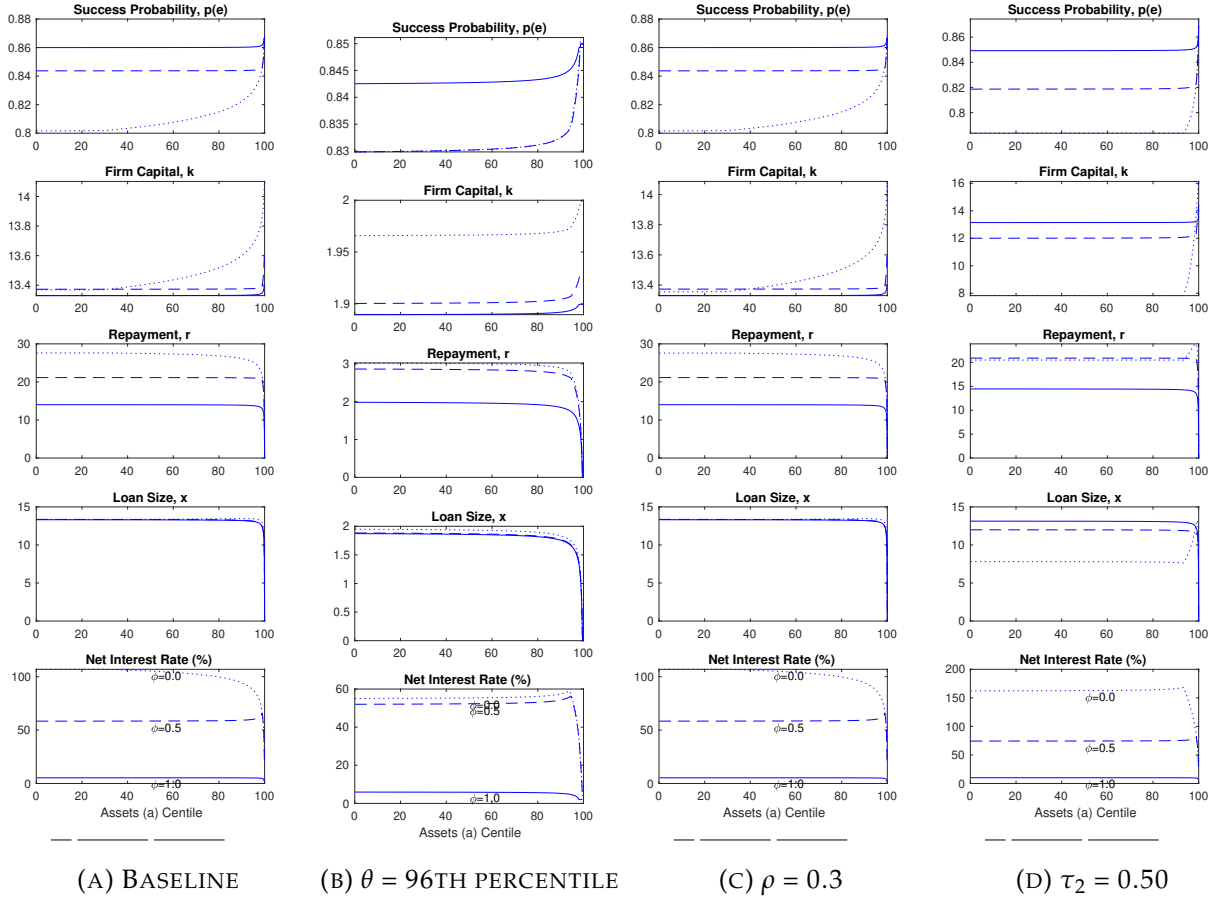
5.1 Credit Contracts

Baseline We begin by looking at credit contracts and capital allocation and how these vary with an entrepreneur's position in the wealth distribution. In all cases, we take a highly productive entrepreneur (at the 99.5th percent of the productivity distribution). Since around 7% of the population are entrepreneurs when there is full access to credit markets, this constitutes the top 7% in the distribution of entrepreneurial productivity and corresponds to a firm size of around 25 employees. This may still seem quite small. However, the firm-size distribution implied by the calibrated values is highly skewed. Such individuals are always active as

---

[26]This is the same as the rate assumed by Paulson et al (2006).

entrepreneurs in our calibration even if they have little wealth and hence we do not need to worry about their occupational choice as we vary parameter values. In all cases, Figure 1 illustrates the outcome for different values of the parameter ϕ. Recall that $\phi = 0$ is the lowest level of competition and $\phi = 1$ is the highest. We also give the contracts for a middle level of competition: $\phi = 1/2$, half the surplus goes to the entrepreneur and half to the borrower. Notice that throughout the horizontal axis refers to centiles of the wealth distribution, not absolute levels of wealth. Since the wealth distribution is highly skewed, absolute differences in wealth are fairly small for much of the lower end of the wealth distribution.

FIGURE 1: EQUILIBRIUM CREDIT CONTRACTS

(A) BASELINE   (B) θ = 96TH PERCENTILE   (C) $\rho = 0.3$ (D) $\tau_2 = 0.50$

Notes: The figures describe characteristics of the equilibrium credit contracts for individuals across all centiles of the asset distribution, holding constant θ. Subfigure (A) presents the baseline scenario where we impose perfect collateralisability of wealth ($\tau_2 = 1$), assume that the distribution of asset holdings and productivities are independent ($\rho = 0$) and presents results for the 99.5th centile of the productivity distribution. Subfigures (B), (C) and (D) preserve the baseline scenario with one exception each: In subfigure (B) we presents results for the 96th centile of the productivity distribution, in subfigure (C) we assume that the distribution of asset holdings and productivities are correlated ($\rho = 0.3$) and in subfigure (D) we impose imperfect collateralisability of wealth ($\tau_2 = 0.5$). Firm outcomes are shown for three distinct levels of competitiveness of credit markets: full competition ($\phi = 1.0$), monopolistic competition ($\phi = 0.0$), and an intermediate level ($\phi = 0.5$). Credit contracts are characterised by the gross interest payment r and the loan size x. Both are measured in absolute terms and in units of the annual income of a US wage laborer (taken to be 43k US$). The net interest rate is calculated as $(r/x - 1) \times 100$. Together with the entrepreneurs own capital x determines the firm capital k. The level of r and x also determine through the effort level the projects success probability $g(e)$.

When interpreting the figures that follow, it should be borne in mind that there are two effects of changing the level of competition. The first is a direct effect whereby the entrepreneur's share of the surplus varies. This affects the total amount of surplus to the extent that incentives for managerial input vary. The second is a general equilibrium effect of competition. Changing competition affects aggregate labour demand, L_D and hence wages thus also changing entrepreneurial profits.

In the top panel of Figure 1, the default probability which depends on managerial input

is illustrated for different wealth levels and competition. Two things are immediate. First, the default rates implied by the model are around 15% across the wealth distribution. This reflects the fact that, in general equilibrium, even low wealth entrepreneurs face good outside options (even when competition is low). This is because, for marginal entrepreneurs, this is the option of being a wage labourer and for higher wealth individuals this is the possibility of self-finance. The default rate is flat across most of the wealth distribution but then decreases for very high wealth individuals who are closer to first-best self-financing. Competition does have real effects since default is lower when there is more competition. This is because the payoffs to entrepreneurs from being successful are higher when there is greater competition.

The second panel gives the use of capital as a function of the position in the wealth distribution. The pattern of default across wealth levels is the flip-side of the repayment probability as illustrated in the top panel. A higher repayment rate naturally means more capital as the marginal product of capital will be higher. Firm capital (k) is lower for lower levels of assets.

The effect of competition on firm capital is non-monotonic for many wealth levels. Moving competition from $\phi = 0$ to $\phi = 1/2$, increases firm capital for individuals at most percentiles of the wealth distribution. However, capital usage typically decreases for the move from $\phi = 1/2$ to $\phi = 1$. At the highest centiles of the wealth distribution increased credit market competition leads monotonically to a decrease in capital usage. Increased capital market competition has two effects on firm capital: it increases capital access, and in general equilibrium it also increases wages. Since labour and capital are complementary in our setting the latter effect depresses capital usage. For some entrepreneurs this latter effect more than offsets the positive effect of increases capital access. For entrepreneurs with high wealth levels the increased capital market competition does not lead to substantially improved credit access, and the positive effect of credit market competition on wages depresses capital usage. Thus the model predicts a heterogeneous, non-linear and often non-monotonic effect of competition on capital allocation which could only be seen by disaggregating by wealth level. That said, the magnitude of these effects is relatively modest, i.e. around a 5% decrease in the amount borrowed for high wealth individuals when competition moves from $\phi = 0$ to $\phi = 0.5$.

The third panel gives the amount repaid by the entrepreneur for the loan that she takes out and the fourth panel gives the loan size. The latter shows that the amount borrowed is almost unaffected by competition. Moreover, the amount borrowed does not depend much on wealth except at very high wealth levels where self-financing substitutes for credit. This pattern reflects the fact the marginal product of capital does not move much with assets in our calibrations.

The repayment level varies with the wealth level. High wealth entrepreneurs tend to be offered low interest rates. The lender's ability to capture entrepreneurial surplus is diminished for high wealth investors since they have a very good autarky outside option and marginal returns to capital are lower at high levels of capital. In contrast to the loan size, the repayment level does vary quite a bit with competition and is highest when competition is low. This reflects the division of surplus between the lenders and entrepreneurs. In highly uncompetitive environments a good amount of the profits that entrepreneurs make are captured by investors. The only limit on this process when $\phi = 0$ is the outside option available and/or the possibility that an entrepreneur receives his efficiency utility.

In popular discussion, the interest rate is frequently used as a barometer of credit conditions. In general our model shows that this is a poor sufficient statistic for matters which

should be gauged from capital allocation and surplus sharing. The reason why the interest rate, $(r/x - 1) \times 100$, is a very poor indicator of capital allocation is that both numerator and denominator are functions of the default rate and the underlying source of heterogeneity (a, θ).

However, it is still interesting to see what the model predicts and how well it relates to efficiency and distribution in the credit market. And we know from many studies of developing country credit markets that interest rates charged by monopolistic borrowers can be very high. With very low competition ($\phi = 0$), the model predicts an interest rate of between 50% and 110% for almost all wealth classes.[27] Only for very high wealth individuals is the rate less than this and it falls quite rapidly for the top of the distribution even with low competition which is due to the fact that outside options for such borrowers are very good (if they need credit at all). The interest rate profile for middle levels of competition is also comparatively flat but again turns down for very high wealth levels. For the highest level of competition, the interest rate is consistently quite low. So competition does seem to have a significant bearing on the interest rate offered to borrowers.

Lower Productivity Benchmark   We now consider what happens when we look at a more marginal group of entrepreneurs by focusing on the 96th percentile in the productivity distribution. In our calibration, these are entrepreneurs who employ around three workers and hence are a eighth of average firm size compared to the baseline case where we focused on the 99.5th percentile of the distribution. We will look at how changing this focus affects credit contracts and credit allocation. These results are depicted in subfigure (B) of Figure 1.

Note first the relationship between competition and default probabilities is much less for these marginal entrepreneurs. However, we still see that at high levels of wealth (above the 80th percentile of the distribution), the repayment probability rises quite steeply. Capital allocation is now much lower (due to productivity being lower) but, in common with the baseline, it is very flat during low levels of the wealth distribution. At the highest levels of the asset distribution both the success probability and the firm capital are independent of assets. Here the first best allocation is achieved. Although repayment and loan size are lower, the same broad relationship with wealth and competition is observed as in the higher productivity case. Interest rates are lower for these less productive borrowers, which is driven by the outside option of both wage labour and self-financing being more attractive in relative terms for these borrowers.

Overall, we find a common pattern that, with optimal second best credit contracts, wealth does not have a strong quantitative effect on the allocation of resources over a wide range of wealth levels. This is because, in a general equilibrium setting, the outside option of wage labour and/or the possibility of receiving an efficiency utility level, does most of the work. This is a general lesson from our model and would only be found by taking a general equilibrium perspective which solves explicitly for the outside options that entrepreneurs face.

Higher Correlation Between Productivity and Wealth   In subfigure (C) of Figure 1, we look at what happens when we allow for a stronger correlation between wealth and productivity,

---

[27]In a model with endogenous savings, this would give an incentive for such groups to save. However, the equilibrium default in our model would imply that borrowers would periodically have their wealth eliminated when the business fails so zero/low wealth individuals would remain a constant feature of an economy in a framework like ours.

which increases the density of high wealth/high productivity individuals. The optimal contracts for each (a, θ) are essentially preserved compared to the baseline.

Frictions in Using Assets as Collateral We now consider what happens when $\tau_2 = 0.5$ so that only 50% of inside capital can be used as collateral. In effect, collateralizable capital in the economy is halved. If the source of the friction is imperfect property rights, then this constitutes the de Soto effect in this model. In our model such a decrease in the collateralizability of inside capital implies that defaulting becomes more attractive for the borrower and more costly to the lender. The lender reacts to both the increase in default and decrease in collateral value by increasing the repayment burden, which amplifies any decrease in managerial effort and hence success probability. As a consequence the lender decreases the loan size and hence firm capital k. In Subfigure (D) of Figure 1 we observe the quantitative implications of this mechanism, where the effects on the default, loan size, firm capital, and interest rate are particularly pronounced in uncompetitive lending environments. For lower levels of competition the effects are muted, since the outside option provided by other lenders limits the lender's ability to increase r in his effort to compensate any decrease in the collateral value. In the case of low competition, the firm's capital is higher for very high-wealth borrowers than in the case of perfect collateralizability of wealth depicted in Subfigure (A). This results from the general equilibrium effect of poor lending conditions: wages are depressed and high-wealth individuals will want to run larger firms. The de Soto effect here leads to a redistribution of firm capital from low to high-wealth individuals.
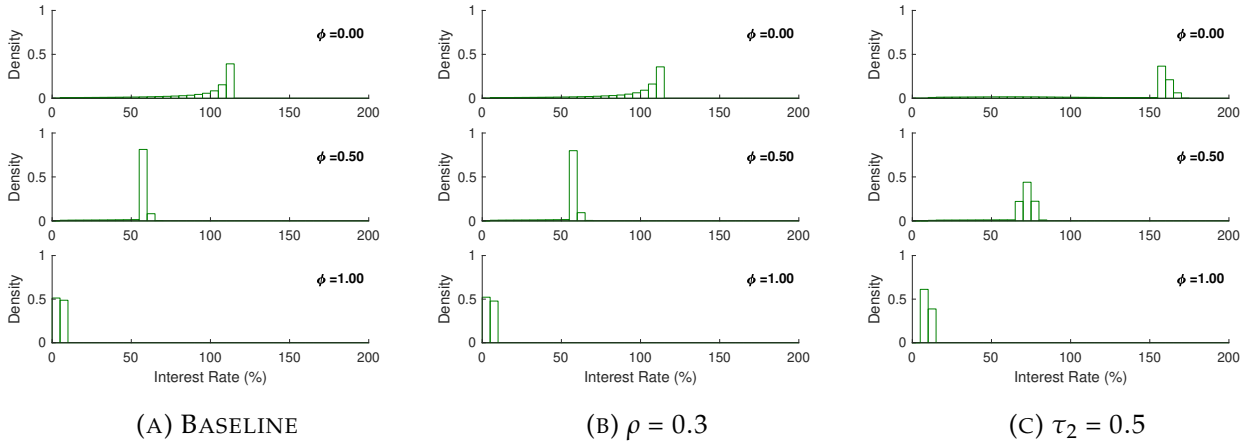
The Distribution of Interest Rates and Default Probabilities Survey evidence that shows high levels and high variance of paid interest rates in developing countries. For example, Aleem (1990) reports a mean interest rate of 78% in his rural lending data with a standard deviation of 38%. Using the Townsend Thai data, Kaboski and Townsend (2012) reports a mean interest rate of 9.5% for 1997-2003 with a standard deviation across borrowers in a year of about 10%. They also report an average default rate of 23%. Variation in the interest rates emerge naturally in our model as a reflection in our framework of heterogeneous repayment rates which reflect underlying differences between borrowers along with market conditions.[28] Here, we will emphasize the importance of competition in affecting the dispersion of interest rates, with dispersion being greatest when competition is low.

Figure 2 looks at the distribution of interest rates across types of borrowers for different levels of competition. As we would expect from Figure 1, when competition is very high then there is no variation in interest rates at all. A feature of the high competition case is the emergence of a modal interest rate; almost every borrower is being offered the same interest rate. The spread of interest rates on offer starts to increase as competition is reduced. However, with $\phi = 0.5$, there is still quite a bit of bunching.

When $\phi = 0$, i.e. no competition, then there is the widest range of interest rates which are essential chosen to extract all of the surplus from a lending relationship between an entrepreneur and lender. This suggests that one factor that could explain the differences in the dispersion of interest rates between Pakistan and Thailand noted above is the competitiveness of credit markets.

[28]Karaivanov and Townsend (2014) also have a moral hazard model which generates such heterogeneity.

(A) BASELINE　　　　　　　　(B) $\rho = 0.3$　　　　　　　　(C) $\tau_2 = 0.5$

Notes: All graphs depict the distribution of interest rates (($r/x - 1) \times 100$) paid by individuals who borrow in equilibrium. Subfigure (A) presents the baseline scenario where we impose perfect collateralisability of wealth ($\tau_2 = 1$), assume that the distribution of asset holdings and productivities are independent ($\rho = 0$). Subfigures (B) and (C) preserve the baseline scenario with one exception each: in subfigure (B) we assume that the distribution of asset holdings and productivities are correlated ($\rho = 0.3$) and in subfigure (C) we impose imperfect collateralisability of wealth ($\tau_2 = 0.5$). For each scenario we show the distribution of interest rates for three distinct levels of competitiveness of credit markets: full competition ($\phi = 1.0$, top figure), monopolistic competition ($\phi = 0.0$, bottom figure), and an intermediate level ($\phi = 0.5$, middle figure).

Figure 2 also allows these distributions to vary across three cases: higher productivity-wealth correlation and worse property rights. The former, as above, leaves things almost unchanged. However, the effect of a lower τ_2, is more visible with higher, and more dispersed interest rates, especially with low credit market competition. This partly reflects that lower collateralizability of wealth leads to higher default rates, and partly that lenders with market power extract surplus through higher interest rates rather than collateral.

FIGURE 3: DISTRIBUTION OF DEFAULT PROBABILITY

(A) BASELINE　　　　　　　　(B) $\rho = 0.3$　　　　　　　　(C) $\tau_2 = 0.5$
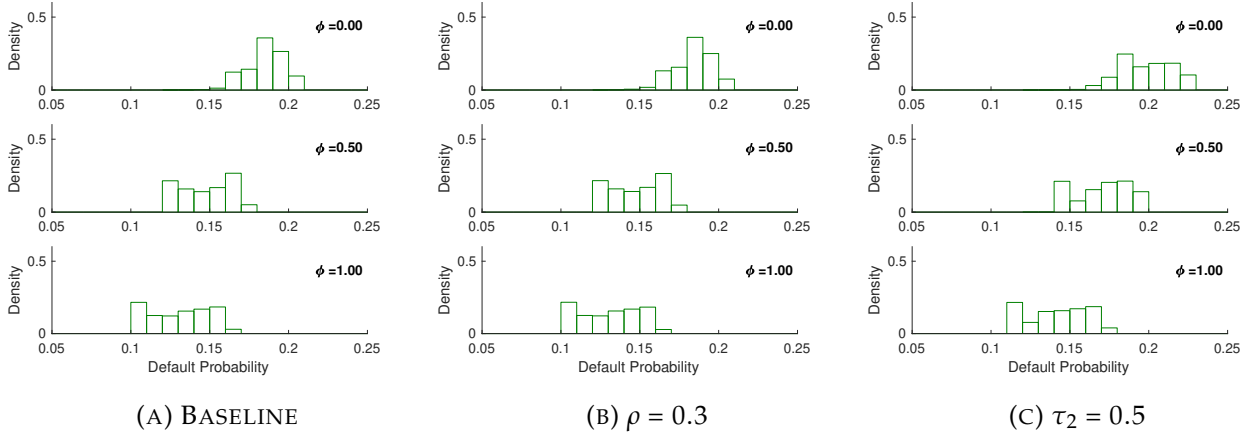
Notes: All graphs depict the distribution of default probabilities ($1 - g(e)$) of individuals who borrow in equilibrium. Subfigure (A) presents the baseline scenario where we impose perfect collateralisability of wealth ($\tau_2 = 1$), assume that the distribution of asset holdings and productivities are independent ($\rho = 0$). Subfigures (B) and (C) preserve the baseline scenario with one exception each: in subfigure (B) we assume that the distribution of asset holdings and productivities are correlated ($\rho = 0.3$) and in subfigure (C) we impose imperfect collateralisability of wealth ($\tau_2 = 0.5$). For each scenario we show the distribution of default probabilities for three distinct levels of competitiveness of credit markets: full competition ($\phi = 1.0$, top figure), monopolistic competition ($\phi = 0.0$, bottom figure), and an intermediate level ($\phi = 0.5$, middle figure).

In Figure 3, we look at the distribution of the credit market distortion across borrower types. A sufficient statistic for this is the default probability $1 - g(e)$. We present figures showing this for different levels of competition. When competition is highest (in the bottom panel), default probabilities are lowest but heterogeneous, with higher wealth borrowers having lower default probabilities. As competition is reduced across the middle and top panels, the distribution

shifts to the right (higher default) as lenders increase the repayment burden r and firm capital k decreases for a broad range of borrowers, both triggering lower levels of managerial effort. These distributions are broadly similar with a change in ρ. Default rates are increasing with decreases in τ_2 as the effective collateral value decreases and hence material effort decreases. This effect is – at least in partial equilibrium – amplified by the responses of the lender namely to increase r and decrease k.

Output   In Figure 4, we look at outcomes among entrepreneurs for the same core parameter values that we looked at in the previous section, i.e., θ in the 99.5th percentile.   We look at output, labour demand, and the income of entrepreneurs. Once again, we look at this in four main cases – the same as those studied when look at credit contracts and capital allocation.

FIGURE 4: EQUILIBRIUM ENTREPRENEURS

(A) BASELINE   (B) $\theta = 96$TH PERCENTILE   (C) $\rho = 0.3$ (D) $\tau_2 = 0.5$
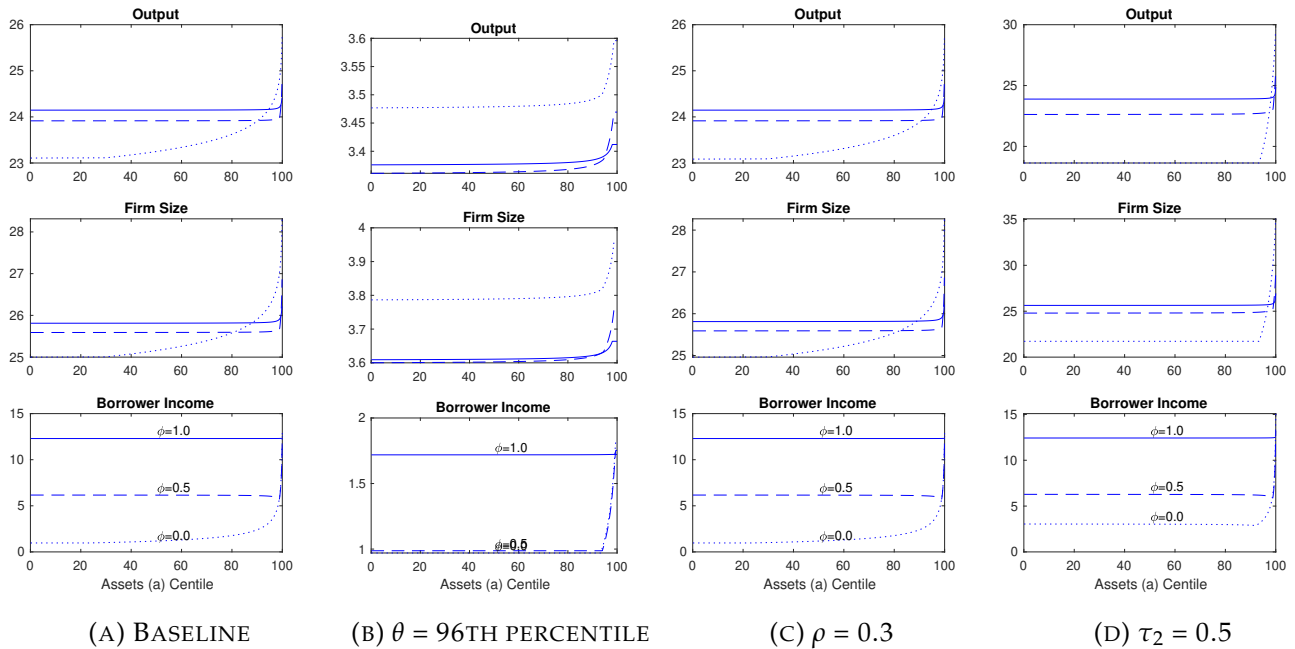
Notes: The figures describe characteristics of the equilibrium firm level outcomes for entrepreneurs across all centiles of the asset distribution, holding constant θ. Subfigure (A) presents the baseline scenario where we impose perfect collateralisability of wealth ($\tau_2 = 1$), assume that the distribution of asset holdings and productivities are independent ($\rho = 0$) and presents results for the 99.5th centile of the productivity distribution. Subfigures (B), (C) and (D) preserve the baseline scenario with one exception each: In subfigure (B) we presents results for the 96th centile of the productivity distribution, in subfigure (C) we assume that the distribution of asset holdings and productivities are correlated ($\rho = 0.3$) and in subfigure (D) we impose imperfect collateralisability of wealth ($\tau_2 = 0.5$). Firm outcomes are shown for three distinct levels of competitiveness of credit markets: full competition ($\phi = 1.0$), monopolistic competition ($\phi = 0.0$), and an intermediate level ($\phi = 0.5$). Expected output is $g(\hat{e})\pi(\hat{k}; \theta)$, measured in units of average annual wage income. Total firm level labor demand is $g(\hat{e})\hat{l} + \hat{e}$, measured in units of one persons annual labor supply. Expected borrower income is $g(\hat{e})(\pi(\hat{k}; \theta) - \hat{r}) + (1 - g(\hat{e}))(\delta\hat{k} - \hat{c}) - w\hat{e}$, measured in units of average annual wage income of labourer in the US.

In the top panel, we look at output (for successful entrepreneurs) for various wealth groups. This shows that output is largely unaffected by wealth up to the 95th percentile of the wealth distribution regardless of the level of competition. However, for the lowest level of competition, output increases from about the 30th percentile onwards. One striking feature of this panel is that for high asset levels, output is highest when competition is lowest. This may seem paradoxical. However, it is worth bearing in mind that the wage increases when competition increases which lowers labour demand and hence production of entrepreneurs. So some entrepreneurs actually produce more when competition is lower. Another feature of this top panel is that once again, the effect of changing competition is highly non-linear with an effect only observable between $\phi = 1/2$ and $\phi = 0$.

The second panel in Figure 4 looks at firm size in terms of its labour force. Again, firms of entrepreneurs with high asset levels are largest when competition is lowest since the wage rate is lowest then. It is clear that taking a general equilibrium model is needed to bring this out. If ϕ could be increased for one entrepreneur rather than for all at once, then the wage would be unchanged and labour demand would fall when competition is decreased. The fact that lack of competition is might lead to an increase in firm size for some entrepreneurs is consistent with the empirical work of Hsieh and Klenow (2009), who find the distribution of firm sizes to be inefficiently concentrated in environments with misallocation of resources. There is no large effect comparing $\phi = 1/2$ and $\phi = 1$.

In the third panel, we look at the income of entrepreneurs. Naturally this depends strongly on the level of competition as the share of the surplus is affected by this. However, it is fairly flat with respect to the level of wealth except when there is very low competition when it turns upwards at the highest wealth levels. This reflects that entrepreneurs with high levels of wealth have the outside option to self-finance.

The remaining panels look at the same variants as in Figure 1. As we might expect, Subfigure (B) shows that having smaller entrepreneurs scales down the size of incomes and labour demand. Competition has a non-monotonic effect on labour demand and income. Increasing the correlation between wealth and productivity again has little effect. Finally, in subfigure (D), the de Soto effect is strongest for low levels of competition, where a decrease in τ_2 leads to markedly lower levels of output and smaller firms. This effect is substantially muted for intermediate or high levels of credit market competition.

As wages rise, there is stronger selection of entrepreneurs from the pool of those who have higher ability which, in turn, leads the size distribution of firms to shift towards larger employers as capital-deepening takes place in the economy, resulting in higher wages. As the financial market access expands further, we get only a small fraction of the population running their own firms with the vast majority relying on supplying labour, but this is good for the workers as wages are higher. While this general equilibrium aspect of the analysis and the labour market channel is in line with the broad thrust of the recent macrodevelopment literature (see Buera et al 2015 for a review), our paper is distinguished by its effort to disentangle the effects of financial access, credit market frictions, and the degree of competition in the credit market.

5.2 Expanding Market Access

In this section, we look at aggregate implications of extending financial inclusion. We consider different values of $z(a, \theta)$ assuming that this is independent of (a, θ) but we allow credit to be extended to a wider and wider set of individuals.

Wages and Self-Employment Figure 5 gives the core aggregate outcomes. The first is the wage as a fraction of the US wage moving form Autarky through to full credit market access. The second gives the fraction of the population that become entrepreneurs. We illustrate this for four cases. The first, the green lines in Figure 5 sets $\phi = 1$ (full competition) and $\tau_2 = 1$ (full collateralizability of wealth) which is the best possible scenario for credit markets. The last, the red line in Figure 5, sets $\phi = 0$ (no competition) and $\tau_2 = 0$ (no collateralizability of wealth). The other lines represent intermediate cases, where either competition is absent but

collaterizability of wealth perfect ($\phi = 0$ and $\tau_2 = 1$, yellow line) or full competition but no collaterizability of wealth ($\phi = 1$ and $\tau_2 = 0$, blue line).

(A) WAGE RATE
(B) ENTREPRENEURSHIP
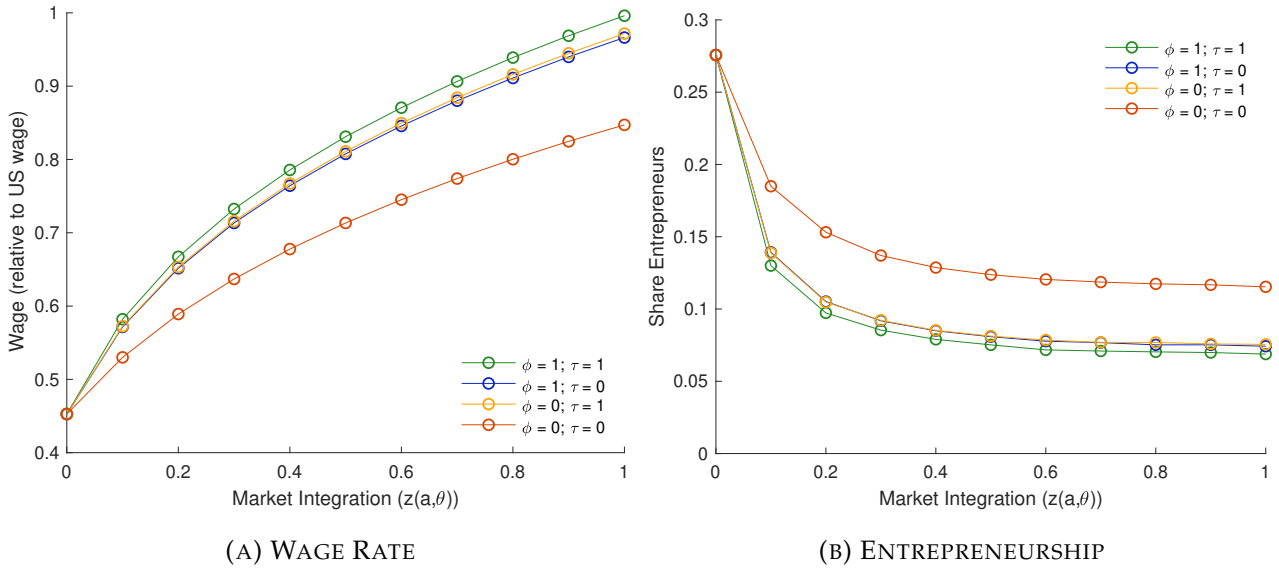
Notes: This figure presents the equilibrium wage rate (Figure A) and the share of entrepreneurs in the population (Figure B) across levels of market integration, ranking from autarky ($z(a, \theta) = 0.0$) to full market integration ($z(a, \theta) = 1.0$). In each figure we present the outcome of interest for the case of perfectly functioning credit markets, subject to credit market access existing ($\phi = 1.0; \tau_2 = 1.0$) and the case of imperfectly functioning credit markets, subject to credit market access existing ($\phi = 0.0; \tau_2 = 0.0$), as well as two intermediate cases. Throughout we assume that the distribution of asset holdings and productivities are independent ($\rho = 0$).

The left hand panel of Figure 5 shows that the wage moves from around 40% of the US wage in autarky to over 80% when there is full credit market access, even with high level of contracting frictions. The upward-sloping curve shows a concave relationship. However, there are still good-sized gains when moving for example, from 40% to 80% access. Comparing the green and red lines, we find that moving from the least to most efficient credit markets (conditional on a level of access) leads to substantial aggregate wage gains although (naturally) the gap between these lines increases as credit market access expands. Interestingly, any one credit market friction alone (lack of competition or imperfect collaterizability of wealth) does not decrease wages substantially. However, the combined incidence of both frictions does have much larger effects on wages. This suggests that either friction amplifies the detrimental effects of the other friction.

The right hand panel in Figure 5 shows the proportion of the population that is an entrepreneur. In autarky this is around 27% of the population. However, it falls rapidly as credit market access expands and then levels off from around 40% access at a little above 7%. Individuals are driven out of self-employment by the increasing wage which makes becoming a wage labourer more attractive and squeezes the profits of marginal entrepreneurs. This figure illustrates why looking at the rate of self-employment is not a good guide to economic outcomes. Individuals are only self-employed because wages are low and they lack access to borrowing opportunities. Thus, allowing for capital to flow to its most productive uses will ensure that only the most productive entrepreneurs (regardless of their wealth) will become entrepreneurs and employment will be concentrated in such firms. Thus the gain are not because the economy has become more intrinsically productive just because gains from trade in labour and capital are realized.

Occupational Choice In Figure 6, we look at the occupational choice at all points in the (a, θ) distribution. The blue shaded area illustrates the space in which individuals choose to be workers and the green shaded area where they choose to be entrepreneurs. Almost all of these entrepreneurs choose to borrow. However, those with very high wealth choose to self-finance.

On the extreme left and right hand panels, we illustrate autarky and the first best. The left hand panel shows, not surprisingly, that the low wealth and low productivity individuals are all workers. However, of the 27% who choose to become entrepreneurs, this goes quite far down the productivity distribution for people with high wealth. In the first best, only around the top 6.8% of the productivity distribution becoming entrepreneurs. Moreover, initial wealth does not matter now since capital allocation in a firm is not dependent on this, only on the productivity of an enterprise.

FIGURE 6: OCCUPATIONAL CHOICE

(A) AUTARKY (B) $z(a,\theta) = 0.10$ (C) $z(a,\theta) = 0.25$ (D) FIRST BEST
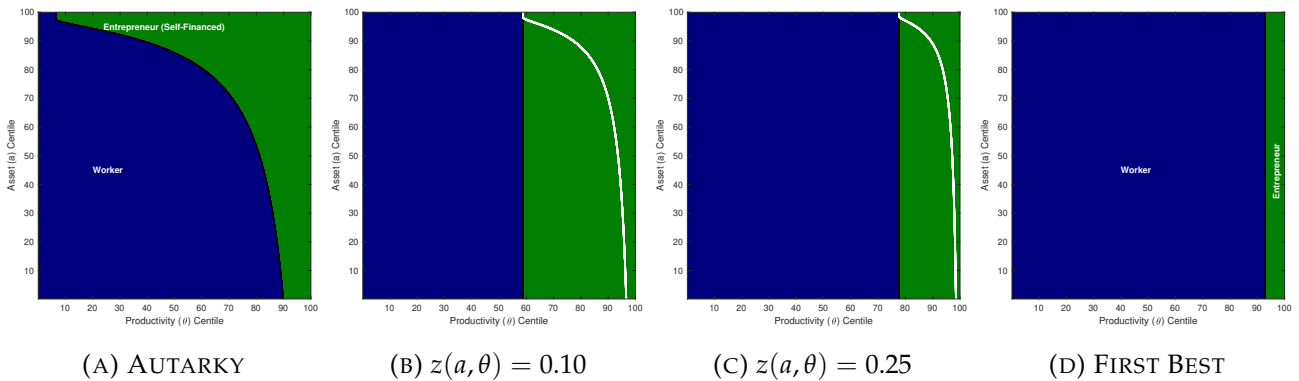
Notes: These figure depicts the occupational choice and lending decision in the a - θ space. Throughout we assume $\rho = 0$. In Figure 6a we present the occupational choice in autarky; in Figure 6b we assume $z(a,\theta) = 0.1$ and $(\phi = 1.0; \tau_2 = 1.0)$; in Figure 6c we assume $z(a,\theta) = 0.25$ and $(\phi = 1.0; \tau_2 = 1.0)$; and in Figure 6d we present results in first best. In Figure 6b and 6c blue and green areas indicate individuals who become workers and entrepreneurs, respectively, when they have credit market access. In those graphs the white line indicates the occupational choice of individuals without credit market access: individuals to the lower-left of it become workers, individuals to the upper-right of it become entrepreneurs.

In the middle two panels, we illustrate the occupational choice for two intermediate values of credit market access where credit contracts are second-best optimal. We choose $z(a,\theta) = 0.10$, i.e. 10% of the population has access to credit markets and $z(a,\theta) = 0.25$ where it is 25%. They show how as credit market access expands, there are fewer entrepreneurs. Even with very limited access, there is switch away from low-productivity high wealth individuals choosing to become entrepreneurs. Hence, the selection effect is quite powerful.[29] The line between the blue and green areas is close to vertical. It then shifts to the right as credit market access expands.

Competition and Distribution We now look at the division of income between labourers, lenders and entrepreneurs. We will consider this as we vary credit market access, $z(a,\theta)$, and the level of competition, ϕ, holding collateralizability of wealth at an intermediate level ($\tau_2 = 0.5$). The size of the columns in Figure 7 illustrate the level of income as a fraction of US income.

On the extreme left is autarky. Here, a little less than half of national income is in the form entrepreneurial profits. The next three bars are for the case where 50% of the population

---

[29]This is similar to the mechanism in Moll (2014) where credit market frictions reduce the wage and induce low productivity individuals to become entrepreneurs.

have access to credit markets but for three levels of competition. With $\phi = 0$, then most of the gains from credit markets are appropriated by lenders and entrepreneurial profits are squeezed relative to autarky. The main effect of increasing competition is to redistribute surplus between lenders and entrepreneurs. There is a modest increase in the wage, due to competition with most of the gain (as we would expect from Figure 5) coming from increasing market access. A similar pattern of surplus redistribution is found for the case of full credit market access.

FIGURE 7: DISTRIBUTION OF SURPLUS ACROSS LEVELS OF COMPETITIVENESS
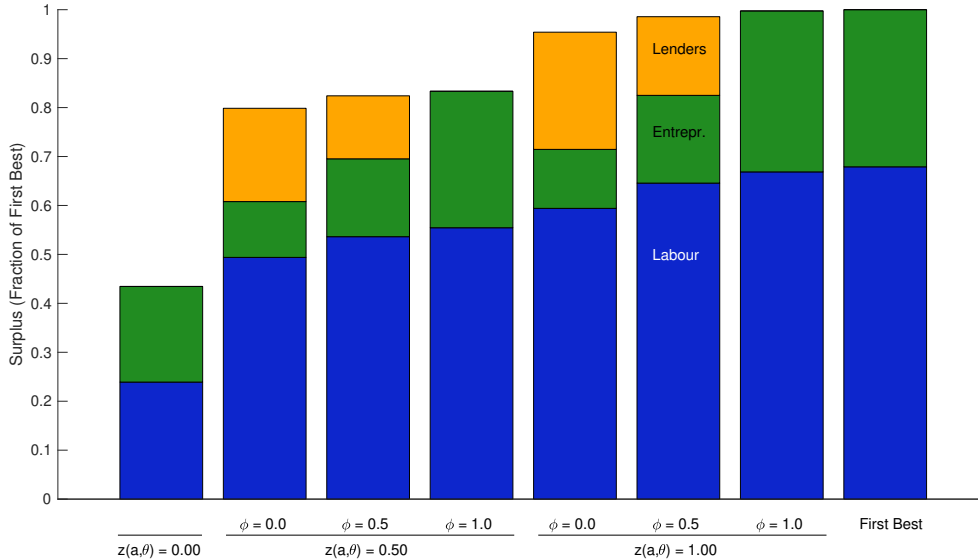
Notes: This figure depicts the size and distribution of total surplus in the economy, in units of first best surplus. It depicts these across levels of market integration, ranking from autarky $(z(a, \theta) = 0.0)$ to full market integration $(z(a, \theta) = 1.0)$, and for distinct levels of competitiveness of credit markets, ranging from monopolistic competition $(\phi = 0.0)$ to full competition $(\phi = 1.0)$. Throughout we impose imperfect collateralizability of wealth $(\tau_2 = 0.5)$ and assume that the distribution of asset holdings and productivities are independent $(\rho = 0)$.

These findings could be relevant for exploring the political economy of credit market expansion and competition. The distributive politics between entrepreneurs and lenders is the most visible dimension of this in Figure 7. But the endogenous wage is also important even though the effects for each worker are small since most of the population are wage laborers; we should therefore expect voting-based politics (assuming that workers understand this) to favor increased competition in credit markets due to the effect on endogenous wages. This is an example, of the kind of factor price effect that has been studied theoretically in the political economy literature, e.g. Acemoglu (2006), but has not been explored quantitatively. Political economy may also reflect the more concentrated interest of entrepreneurs and lenders and will depend on which interest is better organized for lobbying purposes. Credit markets will remain uncompetitive to the extent that lenders are an organized interest.[30]
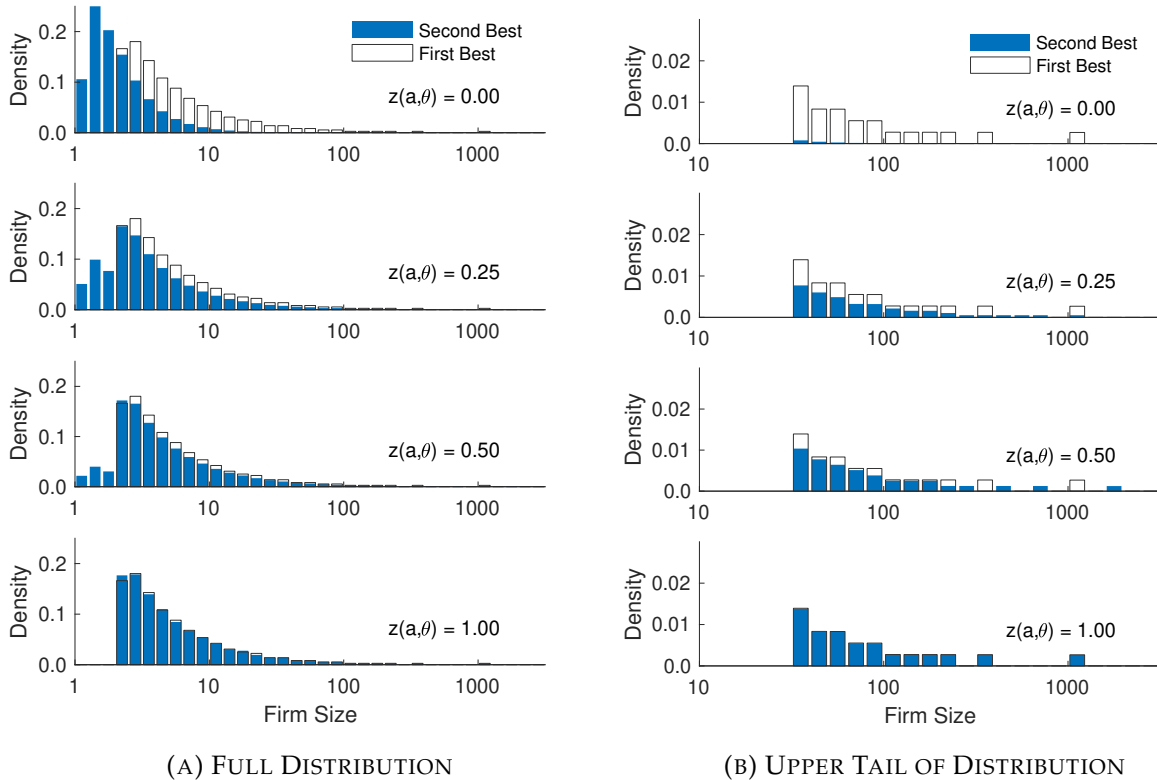
The Distribution of Firm Size   In Figure 8, we look at the distribution of firm size. In each case, we look at variation in market access for four values varying from autarky to full access. Throughout Figure 8, the bars with black outlines give the distribution in the first best and the bars in blue show second best outcomes. The left hand panel gives the full range of firms in the economy whereas the right hand panel gives the distribution of largest firms – the upper

---

[30]Erosa and Hidalgo (2008) consider how entrepreneurs may have a vested interest in poor enforcement to extract rents. However, they do not consider the role of competition affecting the distribution of surplus between lenders and entrepreneurs.

tail of the distribution. In each panel, we give the first best distribution of firms so that we can compare this to the distribution implied by the second-best.

In the top left panel we compare autarky to the first best. Here, we find a very clear shift in the distribution towards small firms. This made even more apparent in the top right-hand panel which shows that there are virtually no large firms in the economy. This lack of labour demand is what keeps the wage low. This broad pattern is found in all of the panels. However, as credit market access varies, the deviation from the "first best" distribution of firm size diminishes. By the time of full credit market access, the first best is very similar to distribution generated by second-best credit markets.

FIGURE 8: DISTRIBUTION OF FIRM SIZES

| (A) FULL DISTRIBUTION | (B) UPPER TAIL OF DISTRIBUTION |

Notes: All graphs depict the distribution of firm sizes in labor units in equilibrium (grey filled bars) We show the distribution of firm sizes for four distinct levels of credit market access: autarky ($z(a,\theta) = 0.0$, top figure), low credit market access ($z(a,\theta) = 0.25$, second from top figure), half the population has credit market access ($z(a,\theta) = 0.5$, third from top figure), and full access ($z(a,\theta) = 1.0$, bottom figure). Throughout we impose perfect collateralisability of wealth ($\tau_2 = 1$), perfect credit market competition ($\phi = 1.0$) assume that the distribution of asset holdings and productivities are independent ($\rho = 0$). Subfigure (A) presents the full distribution of firm sizes with $\log_{10}$ scale on the *x*-axis; subfigure (B) presents a zoomed-in version of the right tail of the firm size distribution. In all figures the distribution of firm sizes in first best is also shown as benchmark (black outlined bars).

The gains from credit market access Figure 9 gives an insight into how the gains from participation in credit markets is distributed across the population. It contains two panels based on the level of competition. Comparing the two panels, it is clear that the gains and losses of credit markets, relative to autarky, are highly heterogeneous. This is important in thinking about possibilities for targeting market access towards particular sub-populations. That said, a key lesson is that distributional outcomes are largely driven by general equilibrium effects on wages and, to that extent, the specific targeting of credit may be less important on the impact on occupational choice and labour demand.

Those who are workers in autarky are all better off with the possibility of trading in credit markets. However, among those who were entrepreneurs in autarky, a large fraction also lose

(A) NO COMPETITION, $\phi = 0$ (B) FULL COMPETITION, $\phi = 1$ (WINZORISED)
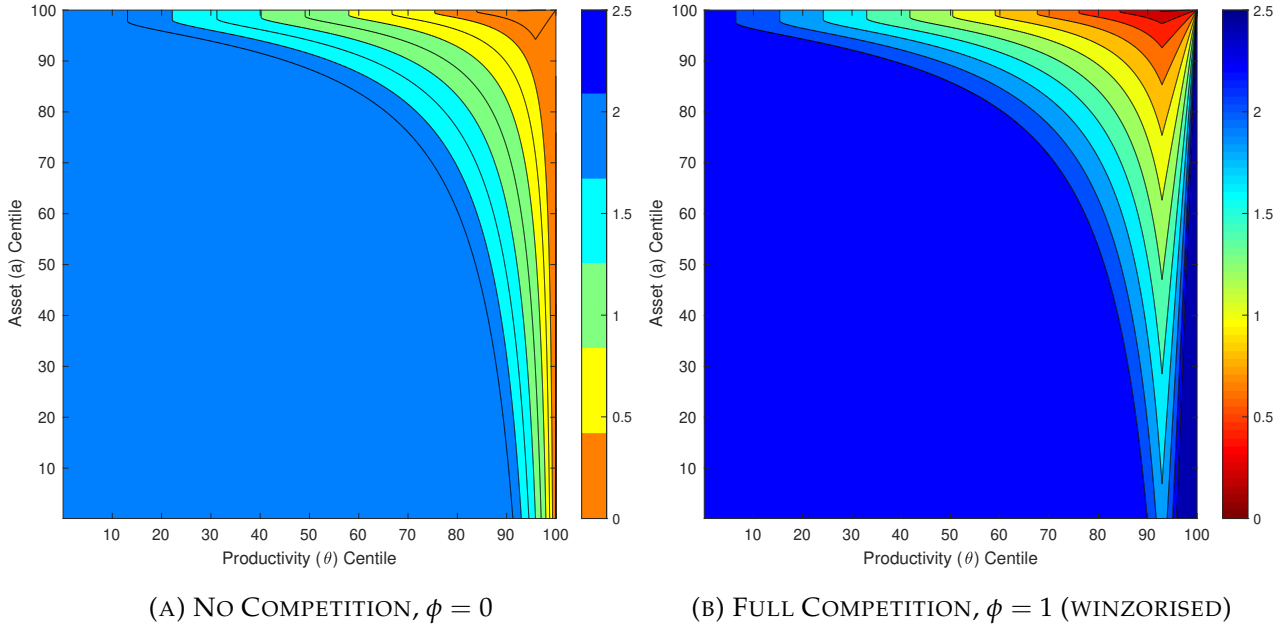
Notes: We calculate for each level of assets a and productivity θ the ratio of income when credit markets exists over income in autarky. Both figures present contour maps of these ratios over the (a, θ) space. Figure (A) assumes fully uncompetitive credit markets ($\phi = 0$) and Figure (B) assumes fully competitive credit markets ($\phi = 1$). Note that in Figure (B) the income ratio is winzorised at 2.5 for visual clarity. This affects the top 3.6% of the θ distribution, and increasingly at lower levels of assets. The highest relative income gain with a ratio of 43.3 is observed for the highest level of θ and lowest level of a. Throughout we assume no correlation between assets and productivity ($\rho = 0$) and perfect collateralisability of wealth ($\tau_2 = 1$).

from the introduction of credit markets due to rising wages. These losses are concentrated amongst entrepreneurs with higher level of wealth, holding θ constant. These entrepreneurs had good access to capital even in autarky, and benefited from low wage levels. This effect is particularly pronounced in the right hand panel of Figure 9, where we compare payoffs with credit markets (with $\phi = 1$) to autarky. Here, high productivity entrepreneurs with low levels of assets benefit form the increased access to credit, despite the fact that this comes with sizeable wage increases. However, entrepreneurs with similarly high productivity and high levels of assets do still loose out relative to autarchy. For them the increased credit access is less important, since they can to a large extent, just self-finance their investment. In the absence of competition, all of the most productive entrepreneurs lose relative to autarchy, since the increase in wages is not compensated by better credit access.

6 Concluding Comments

This paper has explored the role of finance in affecting development looking at both intensive extensive margins. It has developed a framework to explore implications of contracting frictions in the credit market, such as the extent to which credit markets are competitive, as well as the impact of expanding access to finance. While there are now many papers that have looked at the quantitative importance of frictions in credit markets, we are not aware of such models being used to look at financial inclusion.

In our quantitative assessment, it is inclusion that matters most from a quantitative point of view. Expanding market access changes the economy through a general equilibrium channel whereby credit market leads to better selection of entrepreneurs based on their talent and increases firm size, resulting in greater aggregate labour demand and a higher wage. Given that

most of the poor in developing countries are dependent on wage labour, the biggest effects on poverty reduction from changes in the financial sector come through the transformational effect on occupational choice and rising wages.

The paper also reinforces the message in Moll (2014) who has highlighted the importance of diminishing the extent of small-scale self-employment as development progresses with such individuals mainly switching to wage labour. This is consistent with explorations of the impact of micro-credit such as Augsberg et al (2015). While entrepreneurship is important to development, it is the capacity of the market system to allocate capital to those entrepreneurs with high productivity that matters the most. When market access is limited, many of them have to rely on their own wealth and resources which implies that they cannot operate at an optimum scale.

The paper has focused on differences in the *level* as opposed to the growth of income. A natural next step in the research agenda is to look at dynamic implications. Once there are new technologies and shocks to individual firm productivity due to this, the economy has to continuously re-allocate capital. Limited market access plays a role in constraining this aspect of resource allocation too if those who currently enjoy market access can get credit and those with new technologies are excluded. This will affect the ability of the economy to benefit from growth opportunities.

It would be interesting to develop a dynamic version of the model with saving and wealth accumulation. Developing this aspect of our model would be interesting but also challenging; having a positive level of business failure among entrepreneurs would add an interesting new dimension since even talented entrepreneurs might sometimes find themselves with zero wealth. So saving could not in our framework fully alleviate credit market imperfections. If there were also aggregate shocks to business conditions which affect default, these could have a persistent effect on economic performance even in the absence of serially correlated technology shocks. Such an effect is likely to be relevant in understanding the experience of low productivity following the recent financial crisis.

There are many other potential avenues for developing the ideas in this paper. One key role of theory is to provide a way of studying heterogeneity across economies and individuals in the impact of increasing access to credit. While we have shown that the gains from market access differ across wealth, productivity and aggregate features of the economy, there is more that can be done to explore heterogeneous returns as a means of informing policy priorities. This for example could inform where to roll our credit programs geographically to extend the outreach of markets. And there may be scope to target specific unbanked populations based on the heterogeneous gains in different environments predicted by a model like ours.

Another interesting direction for work in this area for which a model along the lines developed here could be useful is exploring complementarities between extending credit market access and other things that governments do to raise productivity such as providing infrastructure or increasing human capital. We have also not allowed the extension of financial inclusion to be targeted, assuming that a representative group of individuals gains access when an extension takes place. It would be interesting to consider how targeted access could make a difference.

The framework that we have proposed has put a spotlight on the behavior of lenders. Our focus here has been on competitive conditions in credit markets which does not appear to have received much attention. But it would also be interesting to explore other frictions. One is the

idea that lenders are subject to behavioral biases as in the animal spirits model of Akerlof and Shiller (2009). Irrational exuberance or caution in lenders' estimates of default probabilities could have real effects on the economy. Exploring this quantitatively would be an interesting extension of the framework.

Finally, we have used a specific technology for providing credit where limits on conventional collateral create a friction, conditional on having access to credit. In ongoing work we are exploring the potential gains from expanding collateral to non-pecuniary punishments, often referred to as "social collateral" which are folded into many microcredit programs. This would allow us to link the paper to discussions about the role of microcredit in development. We plan to explore, using the framework developed here, the quantitative implications of such programs and to address the question of why the returns found in such programs are so heterogeneous in randomized interventions.

References

[1] Acemoglu, Daron, (2006), "Modeling Inefficient Institutions," in *Advances in Economic Theory*, Proceedings of 2005 World Congress, Edited by Richard Blundell, Whitney Newey and Torsten Persson, Cambridge University Press, UK, 2006, pp. 341-380.

[2] Acemoglu, Daron and Fabrizio Zilibotti, (1999), "Information Accumulation in Development," *Journal of Economic Growth*, 4(1), 5-38.

[3] Akerlof, George and Robert Shiller, (2009), *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*, Princeton: Princeton University Press.

[4] Aleem, Irfan (1990) "Imperfect Information, Screening, and the Costs of Informal Lending: A Study of a Rural Credit Market in Pakistan. " *The World Bank Economic Review*, 4(3): 329–349.

[5] Augsberg, Britta, Ralph De Haas, Heike Harmgart, Costas Meghir, (2015), "The Impacts of Microcredit: Evidence from Bosnia and Herzegovina." *American Economic Journal: Applied Economics*, 7(1), 183-203.

[6] Axtell, R. L. (2001) "Zipf Distribution of U.S. Firm Sizes" Science, 293(5536), 1818–1820.

[7] Banerjee, Abhijit, and Andrew F. Newman, (1993) "Occupational Choice and the Process of Development." *Journal of Political Economy*, 101 (2): 274–98.

[8] Banerjee, Abhijit V. and Esther Duflo (2005): "Growth Theory through the Lens of Development Economics", Chapter 7 in Philippe Aghion and Steven N. Durlauf (ed.s) *Handbook of Economic Growth*, Volume 1, Part A, 473-552.

[9] Banerjee, Abhijit, and Esther Duflo (2010), "Giving Credit Where it is Due," *Journal of Economic Perspectives*, 24(3), 61-79.

[10] Bardhan, Pranab, (1984), *Land, labour and Rural Poverty*, Columbia University Press, New York and Oxford University Press, New Delhi.

[11] Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta, (2013),"Cross-Country Differences in Productivity: The Role of Allocation and Selection." *American Economic Review*, 103 (1): 305–34.

[12] Besley, Timothy, Konrad Burchardi, and Maitreesh Ghatak (2012), "Incentives and the De Soto Effect,", *The Quarterly Journal of Economics*, 127(1), 237-282.

[13] Besley, Timothy, Isabelle Roland, and John Van Reenen (2020), "The Aggregate Consequences of Default Risk: Evidence from Firm-Level Data e," CEP Discussion Paper No 1672.

[14] Bloom, Nicholas, (2009), "The Impact of Uncertainty Shocks" *Econometrica*, 77(3), 623-685.

[15] Bloom, Nicholas, and John Van Reenen (2010), "Why Do Management Practices Differ across Firms and Countries? " *Journal of Economic Perspectives*, 24(1), 203–224.

[16] Board of Governors of the Federal Reserve (2016) "Charge-Off and Delinquency Rates on Loans and Leases at Commercial Banks," (Online; accessed 18-November-2016)

[17] Buera, Francisco, Joseph Kaboski, and Yongseok Shin, (2011), "Finance and Development: A Tale of Two Sectors." *American Economic Review*, 101 (5): 1964–2002.

[18] Buera, Francisco, Joseph Kaboski, and Yongseok Shin, (2015), "Entrepreneurship and Financial Frictions: A Macro-Development Perspective." *Annual Review of Economics*, 7, 409-436

[19] Buera, Francisco, Joseph Kaboski, and Yongseok Shin, (2017), "The Macroeconomics of Microfinance," available at https://www.dropbox.com/s/22cnqonxdqdrsu6/BueraKaboskiShin2017.pdf?dl=0

[20] Burgess, Robin and Rohini Pande, (2005), "Do Rural Banks Matter? Evidence from the Indian Social Banking Experiment," *American Economic Review*, 95(3), 780-795.

[21] Caselli, Francesco, (2005), "Accounting for Cross-Country Income Differences," Chapter 9 in Philippe Aghion and Steven N. Durlauf (eds), *Handbook of Economic Growth*, Volume 1, Part A, 679-741.

[22] Cihak, Martin, Asli Demirguc -Kunt, Eric Feyen, and Ross Levine, (2013), "Financial Development in 205 Economies, 1960 to 2010," Working paper 18946, National Bureau of Economic Research.

[23] Credit Suisse (2015), "Global Wealth Report," Credit Suisse Research Institute.

[24] Dabla-Norris, Era, Yan Ji, Robert M. Townsend, and D. Filiz Unsal, (2021), "Distinguishing Constraints on Financial Inclusion and Their Impact on GDP, TFP, and the Distribution of Income," *Journal of Monetary Economics*, 117, 1-18.

[25] Demirguc-Kunt, Asli and Leora Klapper, (2013), "Measuring Financial Inclusion: Explaining Variation in Use of Financial Services across and within Countries," *Brookings Papers on Economic Activity* (Spring), 279-327.

[26] Demirguc-Kunt, Asli, Leora Klapper and Dorothe Singer, (2017), "Financial Inclusion and Inclusive Growth: A Review of Recent Empirical Evidence," World Bank Policy Research Working Paper No. 8040.

[27] Dupas, Pascaline, Dean Karlan, Jonathan Robinson and Diego Ubfal, (2017), "Banking the Unbanked? Evidence from three countries," unpublished typescript.

[28] Erosa, Andres and Ana Hidalgo, (2008), "On Finance as a Theory of TFP, Cross-Industry Productivity Differences, and Economic Rents", *International Economic Review*, 49(2), 437–473.

[29] Galor, Oded, and Joseph Zeira, (1993) "Income Distribution and Macroeconomics." *Review of Economic Studies* 60 (1): 35–52.

[30] Gerschenkron, Alexander, (1962), *Economic Backwardness in Historical Perspective – A Book of Essays*, Harvard University Press, Cambridge, MA.

[31] Ghatak, Maitreesh, Massimo Morelli, and Tomas Sjöström, (2007), "Entrepreneurial Talent, Occupational Choice, and Trickle Up Policies." *Journal of Economic Theory*, 137(1), 27-48.

[32] Ghatak, Maitreesh, (2015), "Theories of Poverty Traps and Anti-Poverty Policies." *World Bank Economic Review, Papers and Proceedings of the ABCDE*, Vol 29 (Supplement 1): S77-S105.

[33] Goldsmith, Raymond W. (1969), *Financial Structure and Development*, Yale University Press, New Haven, CT.

[34] Jeong, Hyeok, and Robert M. Townsend, (2007) "Sources of TFP Growth: Occupational Choice and Financial Deepening." *Economic Theory*, 32 (1): 179–221.

[35] Hsieh, Chang-Tai, and Peter J. Klenow, (2009), "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124 (4): 1403–48.

[36] Hsieh, Chang-Tai, and Peter J. Klenow, (2010), "Development Accounting," *American Economic Journal: Macroeconomics*, 2(1), 207-23.

[37] Itskhoki, Oleg and Benjamin Moll, (2019), "Optimal Development Policies with Financial Frictions," *Econometrica*, 87(1), 139-173.

[38] Kaboski, Joseph P., and Robert M. Townsend, (2011) "A Structural Evaluation of A Large-Scale Quasi-Experimental Microfinance Initiative," *Econometrica*, 79(5), 1357-1406.

[39] Kaboski, Joseph P., and Robert M. Townsend, (2012) "The Impact of Credit on Village Economies," *American Economic Journal: Applied Economics*, 4(2), 98-133.

[40] Karaivanov, Alexander and Robert M. Townsend, (2014), "Dynamic Financial Constraints: Distinguishing Mechanism Design From Exogenously Incomplete Regimes," *Econometrica*, 82(3), 887-959,

[41] Karlan, Dean and Jonathan Zinman, (2009), "Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment," *Econometrica*, 77(6), 1993-2008.

[42] Lee, Ahrang, (2011), "Welfare Losses from Financial Frictions: The Role of Fixed Costs," *Working Paper*.

[43] Levine, Ross, (2005) "Finance and Growth: Theory and Evidence." In *Handbook of Economic Growth*, Vol. 1, Part A, edited by Philippe Aghion and Steven Durlauf, 865–934. Amsterdam: Elsevier B.V.

[44] Lloyd-Ellis, Huw, and Dan Bernhardt.(2000): "Enterprise, Inequality and Economic Development." *The Review of Economic Studies*, 67(1), 147–168.

[45] Matsuyama, Kiminori, (2007), "Aggregate Implications of Credit Market Imperfections." in *NBER Macroeconomics Annual* 2007, Vol. 22, edited by Daron Acemoglu, Kenneth Rogoff, and Michael Woodford, 1–60.

[46] Midrigan, Virgiliu, and Daniel Yi Xu (2014), "Finance and Misallocation: Evidence from Plant-Level Data," *American Economic Review* 104 (2), 422-458.

[47] Moll, Benjamin (2014), "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?," *American Economic Review* 104 (10), 3186-3221.

[48] Nelsen, Roger B. (1999), *An Introduction to Copulas*, Springer.

[49] Paulson, Anna L., Robert M. Townsend and Alexander Karaivanov, (2006), "Distinguishing Limited Liability from Moral Hazard in a Model of Entrepreneurship," *Journal of Political Economy*, 114(1), 100-144.

[50] Restuccia, Diego, and Richard Rogerson, (2008) "Policy Distortions and Aggregate Productivity with Heterogeneous Plants." *Review of Economic Dynamics* 11 (4): 707–20.

[51] Sethi, Dinabandhu and Debashis Acharya, (2018), "Financial inclusion and economic growth linkage: some cross country evidence," *Journal of Financial Economic Policy*, 10(3), 369-385.

[52] Sarma, Mandira and Jesim Pais, (2011), "Financial Inclusion and Development," *Journal of International Development*, 23(5), 613-628.

[53] Sklar, Abe, (1959), "Fonctions de répartition à n dimensions et leurs marges", *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.

[54] Stiglitz, Joseph, (1988), "Economic organization, information, and development," Chapter 5 in Hollis B. Chenery and T.N. Srinivasan (eds), *Handbook of Development Economics* Volume 1, 93-160.

[55] Townsend, Robert, (1978), "Intermediation with Costly Bilateral Exchange," *Review of Economic Studies*, 45(3), 417-25.

[56] Townsend, Robert and Kenichi Ueda, (2006), "Financial Deepening, Inequality, and Growth: A Model-Based Quantitative Evaluation," *Review of Economic Studies*, 73(1), 251-93.

[57] Trivedi, Pravin K. and David M. Zimmer (2007), "Copula Modeling: An Introduction for Practitioners," *Foundations and Trends in Econometrics*, 1(1), 1-111.

[58] World Bank, (2014), Global Financial Development Report 2014: Financial Inclusion, Washington DC, World Bank.

[59] Yi, K. , and J. Zhang, (2016), "Real Interest Rates Over the Long Run," Economic Policy Paper, 16(10), Federal Reserve Bank of Minneapolis.

A Proofs

This section provides the proofs of the results presented in the paper. Recall that the expected payoffs of the lender and the entrepreneur are from equations (3) and (5):

$$\Pi = g(e;\theta)r + (1 - g(e;\theta))c - \gamma x$$
$$V = g(e;\theta)\{\pi(x + \psi a; \theta, \mathbf{p}) - r\} + (1 - g(e;\theta))\{\delta(x + \psi a) - c\} - p_m e + \gamma(1 - \psi)a.$$

The relevant constraints are

(i) the participation constraint (9):

$$V(e, \mathbf{t}; a, \theta, \mathbf{p}) \geq u(a, \theta, \mathbf{p})$$

(ii) the incentive-compatibility constraint (6):

$$g_e(e;\theta)[\pi(x + \psi a; \theta, \mathbf{p}) - \delta(x + \psi a) - r + c] = p_m$$

(iii) the limited liability constraints (7) and (8):

$$r \leq \pi(x + \psi a) + \tau_1 \gamma(1 - \psi)a$$
$$c \leq \tau_1 \gamma(1 - \psi)a + \tau_2 \delta(x + \psi a).$$

Notice that $\pi(x + \psi a) > \delta(x + \psi a)$, and so the right hand side of the limited liability constraint with respect to r is higher than that of the one with respect to c. For notational simplicity, we omit throughout the dependence of functions on θ and $\mathbf{p}$.

Proof of Lemma 1.

Step 1. We show that under the optimal contract $r \geq c$.

Proof of Step 1. Suppose instead $r < c$ is optimal. Then consider a small increase in r to $r + dr$ and a small decrease in c to $c + dc$ that keeps the borrower's payoff constant, and hold k constant, so we have

$$g(e)dr + (1 - g(e))dc = 0$$

The limited liability constraints will be satisfied as the one with respect to c is more stringent than the one with respect to r. Also, by construction the participation constraint is still satisfied and so this is a feasible contract. Observe that from the incentive compatibility constraint, effort level e will decrease, i.e. $de < 0$. By the envelope theorem we can ignore the effect of the change of e on borrower payoff. The change in lender's payoff (holding k constant) is

$$g_e(e)(r - c)de + g(e)dr + (1 - g(e))dc = g_e(e)(r - c)de > 0$$

Since $r < c$, $de < 0$ and $g_e(e) > 0$, this implies that the lender is made better off, contradicting the conjecture of optimality.

Step 2. We show that under the optimal contract $\psi = 1$ (and hence $k = x + a$).

Proof of Step 2. Suppose that $\psi < 1$, and consider a change in contract such that $-dx = d(\psi a) > 0$ (so the capital in the project is unchanged, but the lender's funds are 1:1 replaced by the entrepreneurs liquidated assets), collateral c is decreased by $\tau_1 \gamma d(\psi a)$ (the amount that can no longer be seized from assets not invested in the project), and r is adjusted to keep the borrower payoff constant. Notice that the limited liability constraint on c as well as the participation constraint of the borrower are both still satisfied by construction. For the borrower's payoff to be unchanged, it needs to hold that

$$(g_e(e)[\pi(x + \psi a) - r + \tau_1 \gamma (1 - \psi)a - (1 - \tau_2)\delta k] - p_m)de - g(e)dr + [(1 - g(e))\tau_1 \gamma - \gamma]d(\psi a) = 0$$

Using the incentive compatibility constraint to simplify (envelope theorem), we find the change in r that keeps the borrower's payoff from the credit contract unchanged to be:

$$dr = \frac{(1 - g(e))\tau_1 \gamma - \gamma}{g(e)}d(\psi a) < 0.$$

Since $dr < 0$, also the limited liability constraint on r will be satisfied with the updated contract. The effect on effort is then found by totally differentiating the borrowers incentive compatibility constraint:

$$\left[p_m \frac{g_{ee}(e)}{g_e(e)}\right] de - g_e(e)dr - g_e(e)\tau_1 \gamma d(\psi a) = 0.$$

The lenders change in payoff is given by

$$d\Pi = g_e(e)\left[r - c\right]de + g(e)dr + [(1 - g(e))[-\tau_1 \gamma] + \gamma]d(\psi a).$$

Using the above expressions for de and dr we can simplify to:

$$d\Pi = \frac{g_e(e)}{p_m g(e)\epsilon(e)}(1 - \tau_1)(r - c)\gamma d(\psi a).$$

As long as $\tau_1 < 1$, we have $d\Pi > 0$, a contradiction of the premise that the lender's profit was being maximized originally. We therefore conclude that the optimal contract satisfies $\psi = 1$.

Proof of Proposition 1.

The proof of Proposition 1 proceeds through 5 steps, some of which are similar to the proof of Proposition 1 in Besley et al. (2012):

Step 1. It is never optimal for both the participation constraint and limited liability constraint to be slack.

Proof of Step 1. Suppose this was optimal. Then the lender could increase both r and c by the same small amount, keeping k constant. Effort e will be unchanged, the participation constraint will be satisfied, and the lender will be better off, a contradiction.

Step 2 We show that: (i) if $r > c$ under the optimal contract, then $c = \tau_2 \delta (x + a)$; (ii) if $c < \tau_2 \delta (x + a)$ under the optimal contract, then $r = c$ and effort is at the first-best level.

44

Proof of Step 2 (i) Suppose instead $c < \tau_2 \delta (x + a)$ is optimal. Then by increasing c and decreasing r by small amount to keep the borrower's payoff constant, and hold k constant, we have as in Step 1 of Lemma 1, $g(e)dr + (1 - g(e))dc = 0$; effort level will be higher (from the incentive compatibility constraint of the borrower), i.e. $de > 0$. And with $r > c$, the lender's payoff is $g_e(e)(r - c)de + g(e)dr + (1 - g(e))dc = g_e(e)(r - c)de > 0$, which implies that the lender is better off. Hence a contradiction.

(ii) Consider the contrapositive statement of (i): if $c < \tau_2 \delta (x + a)$, so the limited compatibility constraint is not binding, then $r \leq c$. From Step 1 of the proof of Lemma 1 we then have $r = c$. Now the lender's problem is to maximize $c - \gamma(k - a)$ subject to the participation constraint $g(e)(\pi(k) - \delta k) + \delta k - c - p_m e \geq u$, as well as the incentive compatibility constraint $g_e(e)(\pi(k) - \delta k) = p_m$. Then the lender will want to choose c as high as possible, i.e. the participation constraint is binding. Then the first order condition with respect to k is $g(e)(\pi_k(k) - \delta) = \gamma - \delta$. Hence effort level and capital level are all at the first-best.

Step 3 We show that (i) there exists $\underline{u}$ such that for $u \in [0, \underline{u})$, the optimal contract is characterized by the interior solutions $e = e_0$, $k = k_0$, $r = r_0$ and the participation constraint is not binding; (ii) for $u \geq \underline{u}$ the participation constraint is binding.

Proof of Step 3 (i) Suppose the participation constraint is slack. Then the limited liability constraint is binding by Step 1. The lender's maximization problem is

$$\max_{e,k} g(e) \left(\pi(k) - \frac{p_m}{g_e(e)} - \delta k \right) + \tau_2 \delta k - \gamma(k - a).$$

We first show that this problem is globally concave. Denote the maximand as $\Gamma(e, k)$. We can derive $\Gamma_{ee} = g_{ee}(e)(\pi(k) - \delta k) - p_m(g\epsilon)_e(e)$, $\Gamma_{kk} = g(e)\pi_{kk}$, and $\Gamma_{ek} = g_e(e)(\pi_k - \delta)$. Then with $\pi_{kk} < 0$, the fact that $g(e)\epsilon(e)$ is increasing in e, as well as the assumption that the second order condition is satisfied in the first best case (i.e. $g(e)\pi_{kk}g_{ee}(e)(\pi(k) - \delta k) > (g_e(e)(\pi_k - \delta))^2$), we have $\Gamma_{ee}\Gamma_{kk} = g(e)\pi_{kk}g_{ee}(e)(\pi(k) - \delta k) - g(e)\pi_{kk}p_m(g\epsilon)_e(e) > g(e)\pi_{kk}g_{ee}(e)(\pi(k) - \delta k) > (g_e(e)(\pi_k - \delta))^2 = (\Gamma_{ek})^2$. This implies that the problem is globally concave.

The solution is characterized by the first-order conditions

$$
\begin{aligned}
g_e(e_0) [\pi(k_0) - \delta k_0] &= [1 + g(e_0) \epsilon(e_0)] p_m \\
g(e_0) [\pi_k(k_0) - \delta] &= \gamma - \tau_2 \delta
\end{aligned}
$$

By Assumption 2 (iii) and the fact that $g_e(e)(\pi(k) - \delta k) - p_m g(e)\epsilon(e)$ is decreasing in e and $g(e)(\pi_k(k) - \delta)$ is decreasing in k, the unique global maximum (e_0, k_0) is an interior solution.

Using the binding limited liability constraint and the incentive compatibility constraint, we can rewrite the participation constraint in this case as

$$p_m \left(\frac{g(e_0)}{g_e(e_0)} - e_0 \right) + (1 - \tau_2) \delta k_0 \geq u$$

Note that the left hand side is strictly positive by concavity of $g(e)$, and upon partial differentiation, strictly increasing for $e_0, k_0 > 0$. Hence for any pairs of positive e_0 and k_0, we can define a $\underline{u} := p_m \left(\frac{g(e_0)}{g_e(e_0)} - e_0 \right) + (1 - \tau_2) \delta k_0$ such that for any $u \leq \underline{u}$, the participation constraint is indeed slack and the above contract is feasible and optimal.

(ii) Now suppose for any $u \geq \underline{u}$, the participation constraint is not binding. In this scenario the optimal contract is (e_0, k_0) and the limited liability constraint is binding, see Step 1. The borrower's utility from such a contract is $p_m \left(\frac{g(e_0)}{g_e(e_0)} - e_0 \right) + (1 - \tau_2) \delta k_0 = \underline{u}$. The participation constraint not being binding then implies $\underline{u} > u$, a contradiction.

Step 4 We show (i) there exists a $\bar{u}$ such that for all $u \geq \bar{u}$, the first-best (e_{FB}, k_{FB}) is implemented; (ii) $e_0 < e_{FB}$, $k_0 < k_{FB}$ and $\underline{u} < \bar{u}$ and (iii) for any $u \leq \bar{u}$, the optimal contract satisfies $c = \tau_2 \delta k$.

Proof of Step 4 (i) The first-best effort and capital level is characterized in the main body of the paper as e_{FB} and k_{FB}. By Assumption 2 (iii) and the fact that $g(e)\epsilon(e) \geq 0$ and $\tau_2 \geq 0$ we have $\lim_{e \to 0} g_e(e; \theta) (\pi(k; \theta, \mathbf{p}) - \delta k) - p_m \geq \lim_{e \to 0} g_e(e; \theta) (\pi(k; \theta, \mathbf{p}) - \delta k) - (1 + g(e)\epsilon(e))p_m > 0$ for all $k > 0$ and $\lim_{k \to 0} g(e; \theta) (\pi_k(k; \theta, \mathbf{p}) - \delta) > \gamma - \tau_2 \delta \geq \gamma - \delta$ for all $e > 0$, which ensures that the solutions are interior.

Note that $r = c$ is a necessary condition for the first-best to be implemented. Suppose instead $r \neq c$ and yet the first-best is implemented. Then it follows from the incentive compatibility constraint that given k_{FB}, the borrower would not choose $e = e_{FB}$, a contradiction with the first-best being implemented. Now given $r = c$, the lender's optimization problem becomes to maximize $c - \gamma(k_{FB} - a)$ subject to the limited liability constraint $c \leq \tau_2 \delta k_{FB}$ and the participation constraint $g(e_{FB})(\pi(k_{FB}) - \delta k_{FB}) + \delta k_{FB} - p_m e_{FB} - u \geq c$. The lender will want to choose c as high as possible, subject to constraints. Define $\bar{u} := g(e_{FB})(\pi(k_{FB}) - \delta k_{FB}) - p_m e_{FB} + (1 - \tau_2)\delta k_{FB}$, which is the level of u such that both constraints become binding for the same c. Then $u > \bar{u}$ together with the participation constraint imply $\tau_2 \delta k_{FB} > g(e_{FB})(\pi(k_{FB}) - \delta k_{FB}) + \delta k_{FB} - p_m e_{FB} - u \geq c$, i.e. the limited liability constraint is slack, and hence the participation constraint is binding. Given a binding participation constraint, the surplus maximizing (e_{FB}, k_{FB}) is also maximizing lender profits. Hence for $u \geq \bar{u}$, the first-best outcome e_{FB} and k_{FB} is implemented, as long as the lender makes positive profits, with $r = c = g(e_{FB})(\pi(k_{FB}) - \delta k_{FB}) + \delta k_{FB} - u - p_m e_{FB}$.

(ii) Next we show that $e_0 < e_{FB}$ and $k_0 < k_{FB}$. Consider the first-order conditions characterizing (e_0, k_0) and (e_{FB}, k_{FB}). We can write them in the more general form $g_e(e)(\pi(k) - \delta k) = a$ and $g(e)(\pi_k(k) - \delta) = b$. We can then determine how the solutions to the system of two equations change with a and b. We find $\frac{\partial e}{\partial a} = \frac{g \pi_{kk}}{g_{ee}(\pi - \delta k)g\pi_{kk} - (g_e(\pi_k - \delta))^2} < 0$, $\frac{\partial k}{\partial b} = \frac{g_{ee}(\pi - \delta k)}{g_{ee}(\pi - \delta k)g\pi_{kk} - (g_e(\pi_k - \delta))^2} < 0$, and $\frac{\partial e}{\partial b} = \frac{\partial k}{\partial a} = \frac{g_e(\pi_k - \delta)}{(g_e(\pi_k - \delta))^2 - g_{ee}(\pi - \delta)g\pi_{kk}} < 0$. Since $g(e)\epsilon(e) > 0$, it follows that $e_0 < e_{FB}$ and $k_0 < k_{FB}$.

In order to show that $\bar{u} > \underline{u}$ recall that $\underline{u} := p_m(g(e_0)/g_e(e_0) - e_0) + (1 - \tau_2)\delta k_0$. Further we can write $\bar{u} = p_m(g(e_{FB})/g_e(e_{FB}) - e_{FB}) + (1 - \tau_2)\delta k_{FB}$ using the first order condition for the first-best. Since $(g(e)/g_e(e) - e)$ is strictly increasing in e by Assumption 1 (iii), $(1 - \tau_2)\delta k$ is trivially increasing in k, and $e_0 < e_{FB}$ and $k_0 < k_{FB}$ it follows that $\bar{u} > \underline{u}$.

(iii) Suppose the limited liability constraint was not binding for $u \leq \bar{u}$, i.e. $c < \tau_2 \delta k$. Then the participation constraint is binding by Step 1 and by Step 2 (ii) we have $r = c$. Further $u \leq \bar{u}$ implies by definition $u \leq g(e_{FB})(\pi(k_{FB}) - \delta k_{FB}) - p_m e_{FB} + (1 - \tau_2)\delta k_{FB}$ and the participation constraint is $u = g(e_{FB})(\pi(k_{FB}) - \delta k_{FB}) + \delta k_{FB} - c - p_m e_{FB}$. Together they imply $\tau_2 \delta k_{FB} \leq c$, a contradiction.

Step 5 We show for $u \in [\underline{u}, \overline{u})$, (i) the optimal effort $\xi(u)$ and capital level $\varsigma(u)$ are increasing in u, and (ii) $\lim_{u \to \underline{u}} \xi(u) = e_0$, $\lim_{u \to \underline{u}} \varsigma(u) = k_0$, and $\lim_{u \to \overline{u}} \xi(u) = e_{FB}$, $\lim_{u \to \overline{u}} \varsigma(u) = k_{FB}$.

Proof of Step 5 (i) We know that for $u \in [\underline{u}, \overline{u})$ the limited liability constraint with respect to c is binding (from Step 4 (iii)) and the participation constraint is binding (from Step 3 (ii)). We can use the binding limited liability constraint to substitute for c everywhere, the incentive compatibility constraint to substitute for r in the lender's profit function as well as in the binding participation constraint. This reduces the problem to a maximization of lender profits over two dimensions, e and k, and subject to the participation constraint. Write the Lagrangian as:

$$\mathcal{L}(e,x) = g(e)\left[\pi(k) - \delta k - \frac{p_m}{g_e(e)}\right] + \tau_2 \delta k - \gamma(k-a) - \lambda \left\{ g(e)\left[\frac{p_m}{g_e(e)}\right] + (1-\tau_2)\delta k - p_m e - u \right\} \tag{26}$$

The first-order conditions are:

$$g_e(e)[\pi(k) - \delta k] - p_m[1 + g(e)\epsilon(e)] - \lambda g(e)\epsilon(e)p_m = 0 \tag{27}$$

$$g(e)[\pi_k(k) - \delta] + \tau_2\delta - \gamma - \lambda(1-\tau_2)\delta = 0 \tag{28}$$

$$g(e)\left[\frac{p_m}{g_e(e)}\right] + (1-\tau_2)\delta k - p_m e - u = 0. \tag{29}$$

We can rewrite (27) and (28) to obtain:

$$(1-\tau_2)\delta[g_e(e)(\pi(k) - \delta k) - p_m] - p_m g(e)\epsilon(e)[g(e)(\pi_k(k) - \delta) + \delta - \gamma] = 0. \tag{30}$$

Denoting the left hand side as $h(e,k)$, the total differential is $h_e(e,k)de + h_k(e,k)dk = 0$, with

$$
\begin{aligned}
h_e(e,k) &= (1-\tau_2)\delta g_{ee}(e)(\pi(k) - \delta k) - p_m[g(e)\epsilon(e)]_e(g(e)(\pi_k(k) - \delta) + \delta - \gamma) \\
&\quad - p_m g(e)\epsilon(e)g_e(e)(\pi_k(k) - \delta) \\
h_k(e,k) &= (1-\tau_2)\delta g_e(e)(\pi_k(k) - \delta) - p_m g(e)^2 \epsilon(e)\pi_{kk}(k).
\end{aligned}
$$

Notice that $g(e)$ and $\epsilon(e)$ are both positive and increasing in e; that $g_{ee} < 0$ and $\pi_{kk} < 0$ by concavity; that $\pi(k) - \delta k > 0$ from the incentive compatibility constraint together with $r \geq c$ (from Lemma 1); that $\pi_k(k) - \delta$ from the last result and the assumption $\gamma > \delta$. Further from the borrower's incentive compatibility constraint we have $g_e(e)(\pi(k) - \delta k) - p_m = g_e(e)(r - c) \geq 0$ where the latter inequality follows from Step 1 of Lemma 1. By (30) this implies $g(e)(\pi_k(k) - \delta) + \delta - \gamma \geq 0$. Combined these imply $h_e(e,k) < 0$ and $h_k(e,k) > 0$, and $\frac{dk}{de} > 0$.

Total differentiating (29), we get $p_m g(e)\epsilon(e)de + (1-\tau_2)\delta dk - du = 0$, which together with the previous result implies

$$\frac{de}{du} = \left[p_m g(e)\epsilon(e) - (1-\tau_2)\delta \frac{h_e}{h_k}\right]^{-1} > 0. \tag{31}$$

Together with $\frac{dk}{de} > 0$ this immediately also implies $\frac{dk}{du} > 0$. Hence we can write the optimal effort and capital both as increasing function of u, i.e. $\xi(u)$ and $\varsigma(u)$ respectively.

(ii) We shall prove the result by contradiction. Suppose that $\lim_{u \to \underline{u}} \xi(u) > e_0$. For any $\epsilon > 0$, this is equivalent to $\lim_{\epsilon \to 0} \xi(\underline{u} + \epsilon) > e_0$. Denote $\tilde{e} := \lim_{\epsilon \to 0} \xi(\underline{u} + \epsilon)$ and $\tilde{k} := \lim_{\epsilon \to 0} \varsigma(\underline{u} + \epsilon)$. First observe that (30) holds for any $(\xi(u), \varsigma(u))$ with $u \in [\underline{u}, \overline{u})$ and hence

$\frac{d\varsigma(u)}{d\xi(u)} > 0$ holds too. Therefore $\tilde{e} > e_0$ implies $\tilde{k} > k_0$. Further the participation constraint is binding as long as $u \geq \underline{u}$, and hence in particular for any small ϵ:

$$p_m \left[\frac{g\left(\xi(\underline{u} + \epsilon)\right)}{g_e\left(\xi(\underline{u} + \epsilon)\right)} - \xi(\underline{u} + \epsilon) \right] + (1 - \tau_2)\delta\varsigma(\underline{u} + \epsilon) = \underline{u} + \epsilon.$$

Taking limits on both sides as ϵ goes to 0 we obtain:

$$\left[\frac{g\left(\tilde{e}\right)}{g_e\left(\tilde{e}\right)} - \tilde{e} \right] + (1 - \tau_2)\delta\tilde{k} = \underline{u}. \tag{32}$$

Recall that by the definition we have $\underline{u} = p_m \left(\frac{g(e_0)}{g_e(e_0)} - e_0 \right) + (1 - \tau_2)\delta k_0$. But since $\left[\frac{g(e)}{g_e(e)} - e \right]$ is increasing in e and $(1 - \tau_2)\delta k$ is trivially increasing in k, $\tilde{e} > e_0$ and $\tilde{k} > k_0$ implies that the left hand side is larger than the right hand side, a contradiction. An exactly analogous argument can be constructed for $\lim_{u \to \underline{u}} \xi(u) < e_0$ and $\lim_{u \to \underline{u}} \varsigma(u) \neq k_0$.

The argument for $\lim_{u \to \overline{u}} \xi(u) = e_{FB}$ and $\lim_{u \to \overline{u}} \varsigma(u) = k_{FB}$ follows the same steps, and noting that we can write $\overline{u} = p_m \left(\frac{g(e_{FB})}{g_e(e_{FB})} - e_{FB} \right) + (1 - \tau_2)\delta k_{FB}$ by the definition of $\overline{u}$, the incentive compatibility constraint, as well as the fact that $r = c$ for $u = \overline{u}$.

Proof of Corollary 1.

First notice that $\frac{\partial S(\hat{e}, \hat{k}; \theta, \mathbf{p})}{\partial u} = \frac{\partial S(\hat{e}, \hat{k}; \theta, \mathbf{p})}{\partial e} \frac{\partial \hat{e}}{\partial u} + \frac{\partial S(\hat{e}, \hat{k}; \theta, \mathbf{p})}{\partial \hat{k}} \frac{\partial \hat{k}}{\partial u}$. Recall from the proof of Proposition 1 that $(e, k) = (e_0, k_0)$ for any $u \leq \underline{u}$ and $(e, k) = (e_{FB}, k_{FB})$ for any $u \geq \overline{u}$, and hence $\frac{\partial \hat{e}}{\partial u} = \frac{\partial \hat{k}}{\partial u} = \frac{\partial S(\hat{e}, \hat{k}; \theta, \mathbf{p})}{\partial u} = 0$ in both cases. Further, since $S(e, k; \theta, \mathbf{p}) = g(e; \theta) \pi(k; \theta, \mathbf{p}) + [1 - g(e; \theta)] \delta k - \gamma k - p_m e$, we have:

$$S_e(e, k; \theta, \mathbf{p}) = g_e(e; \theta)(\pi(k; \theta, \mathbf{p}) - \delta k) - p_m \tag{33}$$
$$S_k(e, k; \theta, \mathbf{p}) = g(e; \theta)(\pi_k(k; \theta, \mathbf{p}) - \delta) + \delta - \gamma. \tag{34}$$

For any u such that $\underline{u} < u < \overline{u}$ we have $S_e(\hat{e}, \hat{k}; \theta, \mathbf{p}) \geq 0$ and $S_e(\hat{e}, \hat{k}; \theta, \mathbf{p}) \geq 0$, as well as $\frac{\partial \hat{k}}{\partial u} > 0$ and $\frac{\partial \hat{e}}{\partial u} > 0$ by Step 5 (i) of the proof of Proposition 1, and hence $\frac{\partial S(\hat{e}, \hat{k}; \theta, \mathbf{p})}{\partial u} \geq 0$.

Proof of Proposition 2.

Recall from the proof of Corollary 1 that for any $u \in (\underline{u}, \overline{u})$ we have $\frac{\partial S(\hat{e}, \hat{k}; \theta, \mathbf{p})}{\partial u} \geq 0$, and $\frac{S(\hat{e}, \hat{k}; \theta, \mathbf{p})}{\partial u} = 0$ otherwise.

Now consider the comparative static on wealth a. Notice that it impacts surplus only through its effect on the outside option $u(a, \theta, \phi) = \max\{V^{self}(a, \theta, \mathbf{p}), \hat{u}(\phi; \theta, \mathbf{p}), p_l + \gamma a\}$. While $\hat{u}(\phi; \theta, \mathbf{p})$ is independent of a, both $V^{self}(a, \theta, \mathbf{p})$ and $p_l + \gamma a$ are trivially increasing in a. Hence $u(a, \theta, \phi)$ is weakly increasing in a, and so is $S(e, k; \theta, \mathbf{p})$.

For comparative statics on ϕ, we differentiate (17) with respect to ϕ to yield

$$\frac{\partial \hat{u}(\phi; \theta, \mathbf{p})}{\partial \phi} = \frac{\hat{S}(\hat{u}(\phi; \theta, \mathbf{p}); \theta, \mathbf{p})}{1 - \phi \hat{S}_u(\hat{u}(\phi; \theta, \mathbf{p}); \theta, \mathbf{p})} \tag{35}$$

Since $\phi \in [0, 1]$, we have $\frac{\partial \hat{u}(\phi; \theta, \mathbf{p})}{\partial \phi} > 0$ if $S_u(\hat{u}(\phi; \theta, \mathbf{p}); \theta, \mathbf{p}) \leq 1$. To see this, note that we could also write the total surplus of a lending relationship as $\hat{S}(\hat{u}; \theta, \mathbf{p}) = \hat{\Pi}(\hat{u}; \theta, \mathbf{p}) + \hat{u}$, where

$\hat{\Pi}(\hat{u}; \theta, \mathbf{p})$ is the lender's profit. Then

$$\hat{S}_u(\hat{u}; \theta, \mathbf{p}) = 1 + \hat{\Pi}_u(\hat{u}; \theta, \mathbf{p}) \leq 1. \tag{36}$$

The inequality follows from the fact that the lender's profit must be non-increasing as the borrower's outside option goes up, i.e. $\hat{\Pi}_u(\hat{u}; \theta, \mathbf{p}) \leq 0$. If this was not true, the lender could offer another contract which makes both him and the borrower better off, contradicting that the lender is maximizing profits.

Last, we find $\hat{S}_\theta$ as

$$\begin{aligned}
\hat{S}_\theta &= g_\theta(\hat{e}; \theta)[\pi(\hat{k}; \theta, \mathbf{p}) - \delta\hat{k}] + g(\hat{e}; \theta)\pi_\theta(\hat{k}; \theta, \mathbf{p}) \\
&\quad + [g_e(\hat{e}; \theta)[\pi(\hat{k}; \theta, \mathbf{p}) - \delta\hat{k}] - p_m]\hat{e}_\theta + [g(\hat{e}; \theta)[\pi_k(\hat{k}; \theta, \mathbf{p}) - \delta] + \delta - \gamma]\hat{k}_\theta.
\end{aligned} \tag{37}$$

The sign of $g_\theta(\hat{e}; \theta)$ and $g_{e\theta}(\hat{e}; \theta)$ are both indeterminate, and examples can be constructed both such that $\hat{S}_\theta > 0$ and $\hat{S}_\theta < 0$. In particular, we can show that for $g(e; \theta) = \theta e^\alpha$, $\hat{S}_\theta > 0$. For $g(e; \theta) = \lambda(\frac{e}{\theta^\mu})^\alpha$, and $\pi(k; \theta) = A\theta^q k^\beta$ with $0 < \alpha, \beta, q < 1$: if $\mu\alpha < q$, then $\hat{S}_\theta > 0$; if instead $\mu\alpha > q$, then $\hat{S}_\theta > 0$ if $\pi_k(k_0; \theta) < \beta C$; $\hat{S}_\theta < 0$ if $\pi_k(k_{FB}; \theta) > \beta C$ and in addition $\pi_k(k_0; \theta) > C$, where $C = \frac{\delta\mu\alpha}{\mu\alpha - q}$ is a constant.
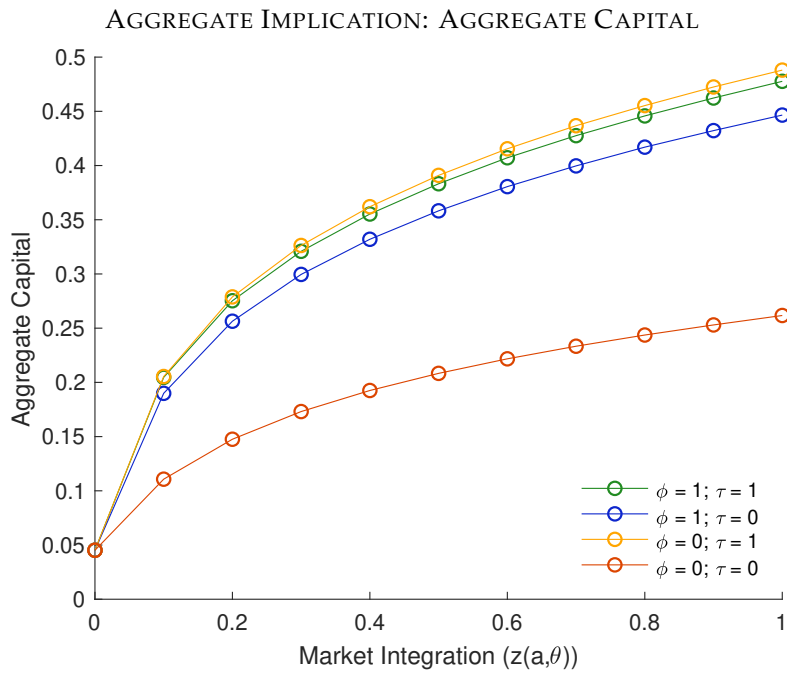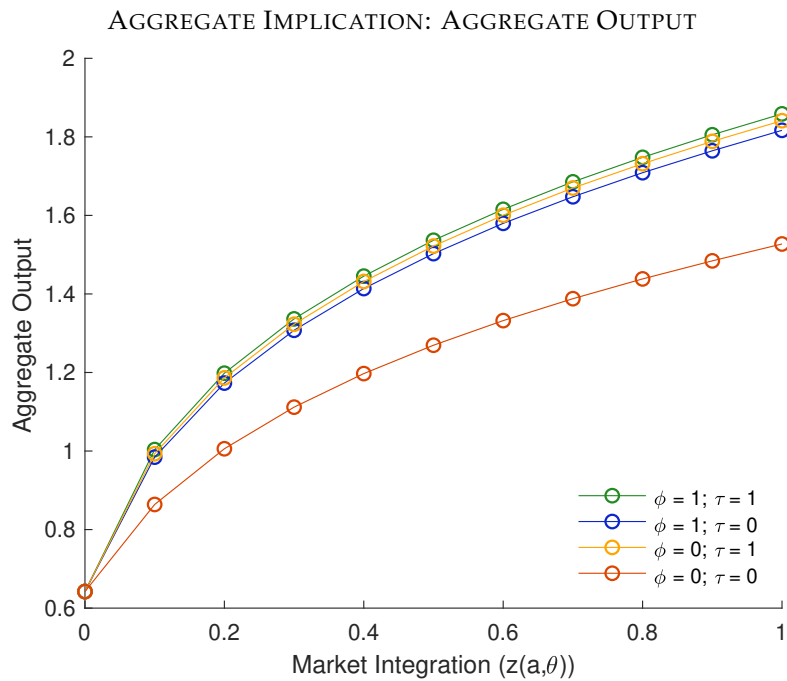
B Additional Figures

AGGREGATE IMPLICATION: AGGREGATE CAPITAL
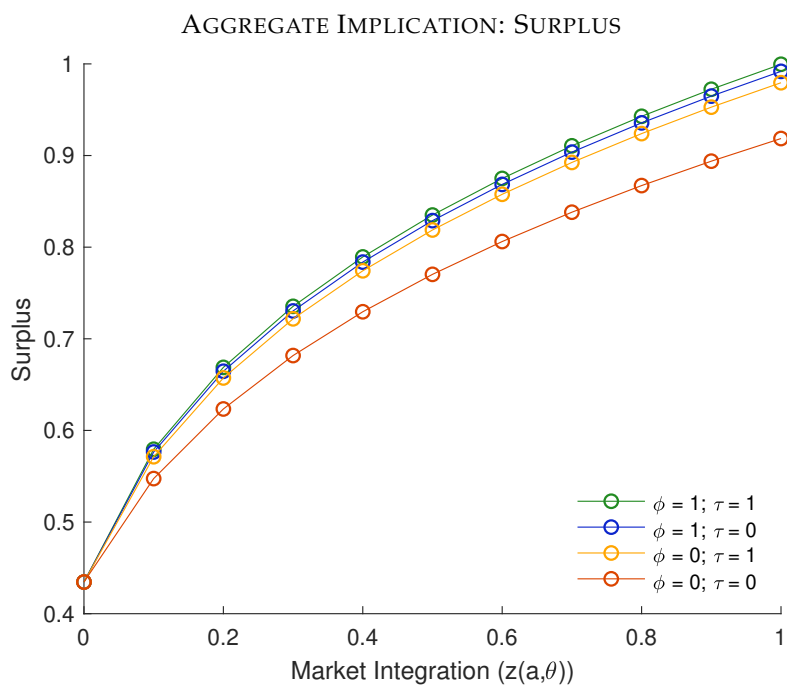

FIGURE 5 (C)

AGGREGATE IMPLICATION: AGGREGATE OUTPUT

FIGURE 5 (D)

AGGREGATE IMPLICATION: SURPLUS

FIGURE 5 (E)

AGGREGATE IMPLICATION: MISALLOCATION

FIGURE 5 (F)

AGGREGATE IMPLICATION: TFP SHORTFALL

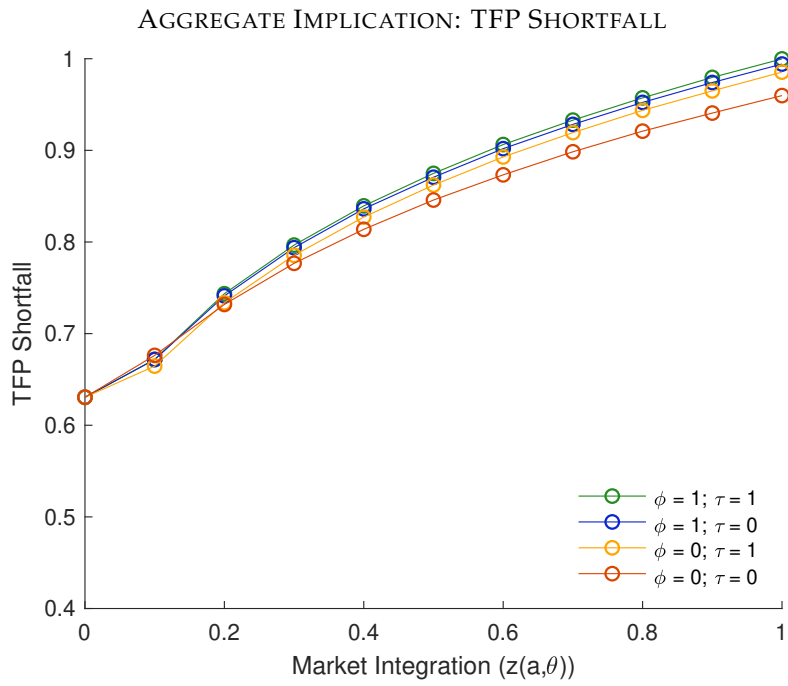FIGURE 5 (G)

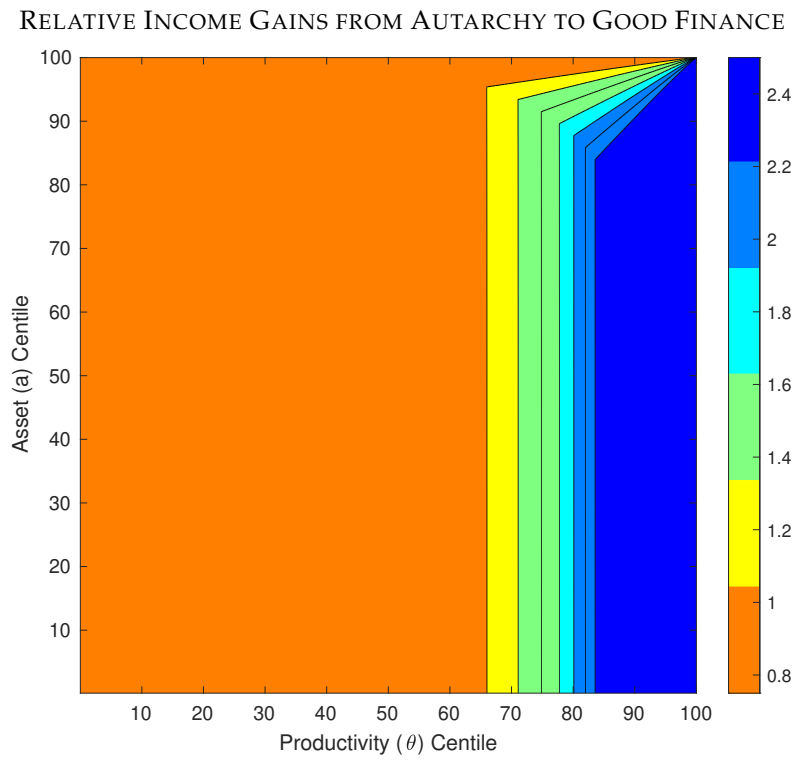RELATIVE INCOME GAINS FROM AUTARCHY TO GOOD FINANCE
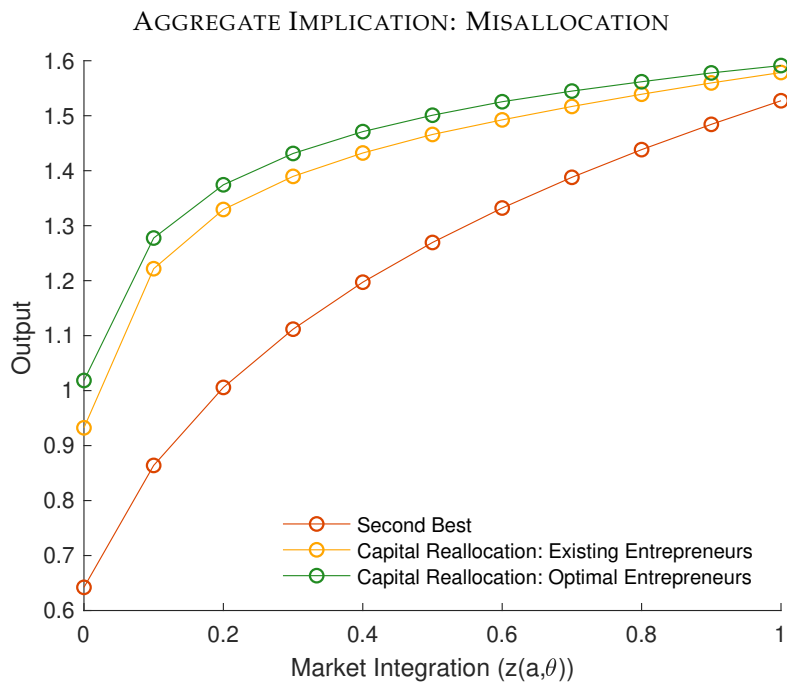
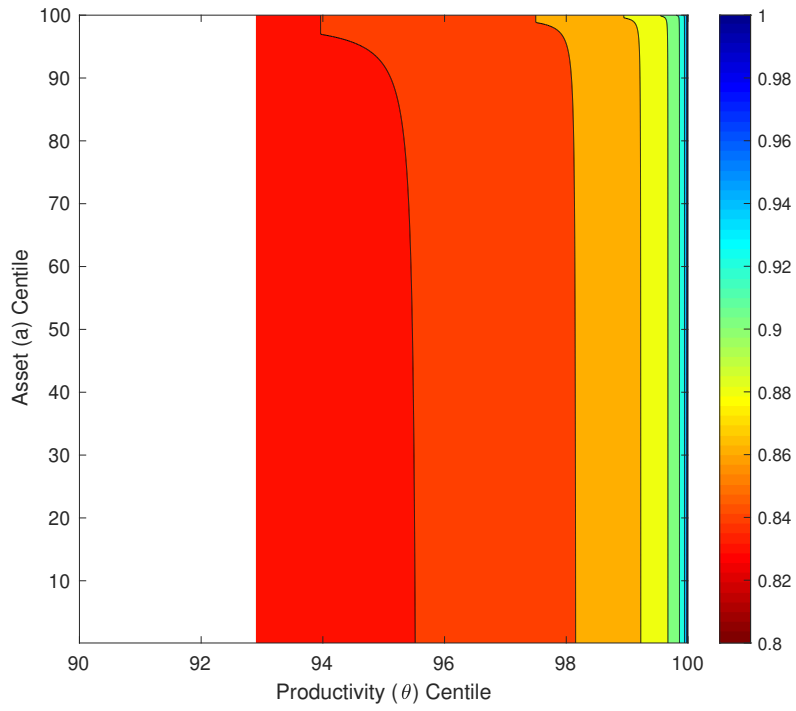FIGURE 9 (C)

COMPETITION AND SUCCESS PROBABILITY

FIGURE 10 (A)

COMPETITION AND SUCCESS PROBABILITY

FIGURE 10 (B)

53