NBER WORKING PAPER SERIES

DISTINGUISHING CONSTRAINTS ON FINANCIAL INCLUSION AND THEIR IMPACT ON GDP, TFP, AND INEQUALITY

Era Dabla-Norris
Yan Ji
Robert M. Townsend
D. Filiz Unsal

Working Paper 20821
http://www.nber.org/papers/w20821

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2015

Previously circulated as "Distinguishing Constraints on Financial Inclusion and Their Impact on GDP and Inequality." This paper is part of a research project on macroeconomic policy in low-income countries supported by the U.K.'s Department of International Development (DFID). This paper should not be reported as representing the views of DFID or the National Bureau of Economic Research. Townsend also acknowledges research funding from the NICHD. We thank Abhijit Banerjee, Adrien Auclert, Francisco Buera, Stijn Claessens, David Marston, Rafael Portillo, Alp Simsek, Iván Werning, and seminar participants in the IMF Workshop on Macroeconomic Policy and Inequality, and the MIT Macro and Development Lunch for very helpful comments. All errors are our own.

Distinguishing Constraints on Financial Inclusion and Their Impact on GDP, TFP, and Inequality
Era Dabla-Norris, Yan Ji, Robert M. Townsend, and D. Filiz Unsal
NBER Working Paper No. 20821
January 2015, Revised June 2015
JEL No. C54,E23,E44,E69,O11,O16,O57

ABSTRACT

We develop a micro-founded general equilibrium model with heterogeneous agents and three dimensions of financial inclusion: access (determined by a participation cost), depth (determined by a borrowing constraint), and intermediation efficiency (determined by a monitoring cost). We find that the economic implications of financial inclusion policies vary with the source of frictions. In partial equilibrium, we show analytically that relaxing each of these constraints separately increases GDP. However, when constraints are relaxed jointly, the impacts on the intensive margin (increasing output per entrepreneur with access to credit) are amplified, while the impacts on the extensive margin (promoting credit access) are dampened. In general equilibrium, we discipline the model with firm-level data from six countries and quantitatively evaluate the policy impacts. Multiple frictions are necessary to match the country-specific variables, e.g., credit access ratio, interest rate spread, and non-performing loans. A TFP decomposition finds that most of the productivity gains are captured by a between-regime shifting effect, whereby talented entrepreneurs obtain credit and expand their businesses. In terms of inequality and welfare, reducing the participation cost benefits talented-but-poor agents the most, while relaxing the borrowing constraint or intermediation cost is more beneficial for talented-and-wealthy agents.

Era Dabla-Norris
International Monetary Fund
700 19th Street Northwest
Washington, DC 20431
EDABLANORRIS@imf.org

Yan Ji
Massachusetts Institute of Technology
77 Massachusetts Avenue E19-750
Cambridge, MA 02139
jiy@mit.edu

Robert M. Townsend
Department of Economics, E17-230
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
rtownsen@mit.edu

D. Filiz Unsal
International Monetary Fund
700 19th Street Northwest
Washington, DC 20431
dunsal@imf.org

1 Introduction

Financial deepening has accelerated in emerging market and low-income countries over the past two decades. The record on financial inclusion, however, has not kept apace. Large amounts of credit do not always correspond to broad use of financial services, as credit is often concentrated among the largest firms. Moreover, firms in developing countries continue to face barriers in accessing financial services. For instance, 95 percent of firms in advanced economies have access to a bank loan or line of credit as compared with 58 percent in developing countries, and 20 percent in low-income countries (Figure 1). Collateral requirements for loans, which impose borrowing constraints on firms, are also two to three times higher in developing countries as compared to advanced economies. Similarly, interest rate spreads (the difference between lending and deposit rates) tend to be much higher than in advanced economies. Firms also differ in terms of their own identification of access to finance as a major obstacle to their operations and growth: in developing countries, 35 percent of small firms report that access to finance is a major obstacle to their operations, compared with 25 percent of large firms, and 8 percent of large firms in advanced economies (Figure 2).[1]

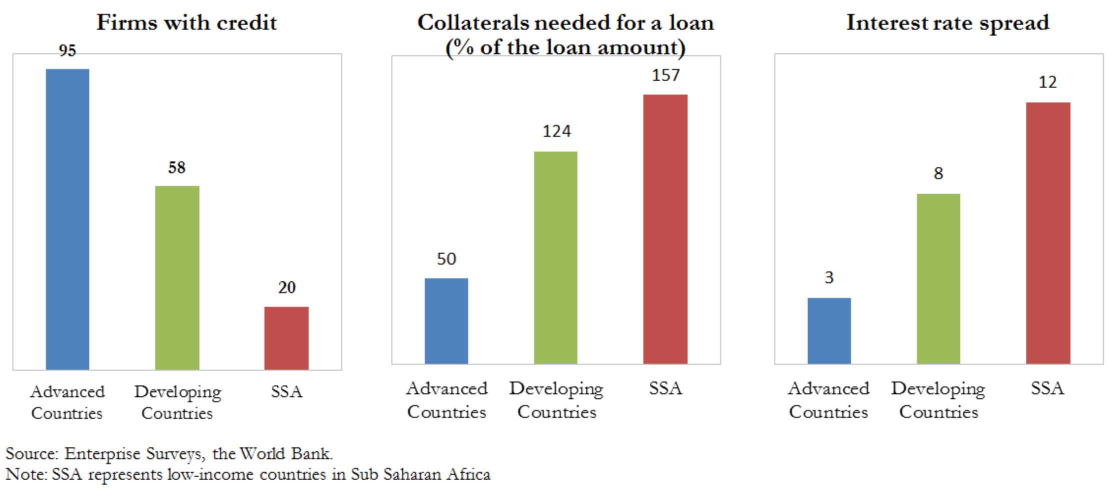
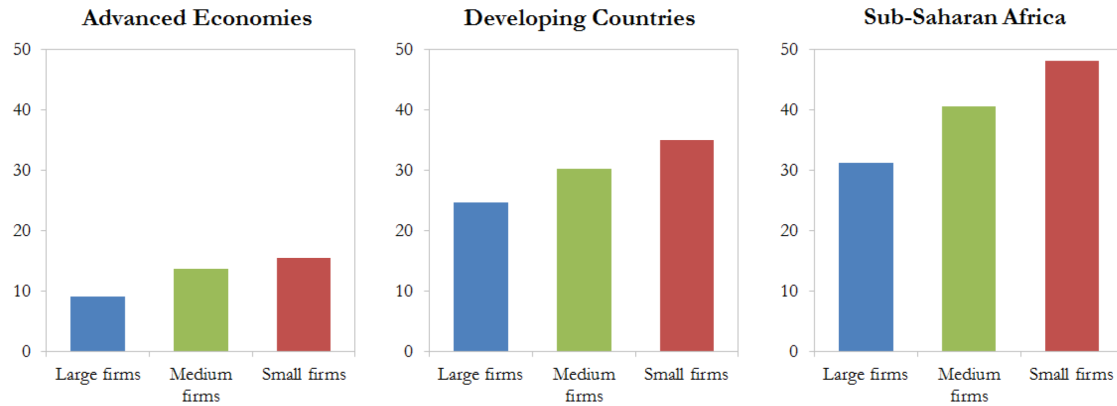


Figure 1: Financial inclusion in the world.

These considerations warrant a tractable framework that allows for a systematic examination of the linkages between financial inclusion, GDP, and inequality. Given that financial inclusion is multi-dimensional, involving both participation barriers and financial frictions that constrain credit availability, policies to foster financial inclusion are likely to vary across countries. In this paper, we develop a micro-founded general equilibrium model to highlight, distinguish, and evaluate the differential impacts of different financial constraints on GDP, TFP, and inequality and examine

[1]This problem is more acute for firms in the informal sector. This paper focuses primarily on firms in the formal sector.



Source: Enterprise Surveys, the World Bank.

Figure 2: Percent of firms identifying access to finance as a major constraint.

how these constraints interact both theoretically and numerically.

In the model, agents are heterogeneous—distinguished from each other by wealth and talent. Agents choose in each period whether to become entrepreneurs or to supply labor for a wage. Workers supply labor to entrepreneurs and are paid the equilibrium wage. Entrepreneurs have access to a technology that uses capital and labor for production. In equilibrium, only talented agents with a certain level of wealth choose to become entrepreneurs. Untalented agents, or those who are talented but wealth constrained, are unable to start a profitable business, choosing instead to become wage earners. Thus, occupational choices determine how agents can save and also what risks they can bear, with long-run implications for growth and the distribution of income.

The model features an economy with two "financial regimes", one with credit and one with savings only. Agents in the savings regime can save (i.e., make a deposit in banks to transfer wealth over time) but cannot borrow. Participation in the savings regime is free, but agents must pay a participation cost to borrow. The size of this participation cost is one of the determinants of financial inclusion, capturing the fixed transactions costs and high annual fees, documentation requirements, and other access barriers facing entrepreneurs in developing countries.

Once in the credit regime, agents can obtain credit, but its size is constrained by two additional types of financial frictions—limited commitment and asymmetric information. These distort the allocation of capital and entrepreneurial talent in the economy, lowering aggregate total factor productivity (TFP). The first financial friction is modeled as a borrowing constraint, which arises from imperfect enforceability of contracts. Entrepreneurs have to post collateral in order to borrow. The value of collateral is thus another determinant of financial inclusion, affecting the amount of credit available. The second financial friction arises from asymmetric information between banks and borrowers. In this environment, interest rates charged on loans must cover the cost of monitoring of highly-leveraged entrepreneurs. Because more productive and poorer agents are more likely to be highly leveraged, the ensuing higher intermediation cost is another source of inefficiency and financial exclusion. As only highly-leveraged entrepreneurs are monitored, entrepreneurs face differential costs of capital and may choose not to borrow even when credit is available.

We distinguish the effect of financial inclusion on the extensive and intensive margins. On the one hand, relaxing financial constraints can increase GDP through the extensive margin by increasing the credit access ratio (i.e., moving entrepreneurs from the savings regime to the credit regime). On the other hand, it enables entrepreneurs in the credit regime to produce more output, which boosts up GDP. This is the effect on the intensive margin.

In a partial equilibrium analysis with fixed interest rates and wages, we show that relaxing the ex-ante friction captured by the credit participation cost and the two ex-post frictions within the credit regime can increase GDP through both the extensive and intensive margins. We obtain closed-form solutions which indicate that relaxing different financial constraints has differential quantitative impacts, depending upon country-specific (the primitive model parameters calibrated from data) and individual-specific (wealth and talent) characteristics. We find that there are non-trivial interactions among the three financial constraints. The credit participation cost, the borrowing constraint, and the intermediation cost have complementary effects on the intensive margin, but are substitutes on the extensive margin. Intuitively, this is because a lower credit participation cost increases the credit access ratio, such that relaxing the borrowing constraint and reducing the intermediation cost have less of an impact. In other words, when the credit participation cost is low, the credit access ratio is already high, so that there is little room for increasing this ratio further through the other two channels. Essentially, the substitution effect on the extensive margin is due to the natural bound on the maximum credit access ratio (100%). On the intensive margin, relaxing one constraints amplifies the effects of relaxing other constraints. This is because, when the credit participation cost is low, entrepreneurs are left with more wealth after entering the credit regime. Since the amount of credit and the total intermediation cost are proportional to wealth, relaxing the borrowing constraint and reducing the intermediation cost increases business profits more.

The general equilibrium effect of financial inclusion does not allow for deriving analytical solutions, since it involves the endogenous distribution of wealth and talent and equilibrium interest rates and wages. To better understand the differential impacts of relaxing the various financial constraints, and in particular, how they interact in general equilibrium, we calibrate the model using data from the World Bank Enterprise Surveys and World Development Indicators. We jointly choose the model's key parameters to match the simulated moments, such as the percent of firms with credit and the firm employment distribution, as well as the economy-wide non-performing loans (NPL) ratio, the interest rate spread, and the bank overhead costs to assets ratio. We calibrate the model separately for six developing countries at varying degrees of economic development:

three low-income countries (Uganda in 2005, Kenya in 2006, and Mozambique in 2006), and three emerging market economies (Malaysia in 2006, the Philippines in 2007, and Egypt in 2007).

The model simulations confirm our partial equilibrium analysis, suggesting that the impact of financial inclusion policies depends upon country-specific characteristics. For example, Uganda's GDP is most responsive to a relaxation of the borrowing constraint. This is because entrepreneurs in Uganda are severely constrained by high collateral requirements, so that reducing the intermediation cost only benefits a small number of highly-leveraged entrepreneurs. By contrast, a high fixed participation cost is a major obstacle to financial inclusion in Malaysia. These results suggest that understanding the specific factors constraining financial inclusion in an economy is critical for tailoring policy advice.

The model simulations also indicate that different dimensions of financial inclusion unambiguously increase the economy's GDP and TFP as talented entrepreneurs, who desire to operate firms at a larger scale, benefit disproportionately. However, they have a differential impact on inequality and there can be trade-offs. For example, a decline in the intermediation cost increases income inequality as it raises the profits of entrepreneurs living in the credit regime (whose income is already higher than others). Relaxing the borrowing constraint, on the other hand, can have an ambiguous impact on inequality, with inequality initially increasing and then declining. In other words, a Kuznets-type response can be generated. In fact, different dimensions of financial inclusion can result in different distributional consequences. In partial equilibrium, everyone can benefit from a more inclusive financial system, albeit to varying degrees. However, in general equilibrium, the resulting changes in interest rates and wages can lead to losses for some agents. For example, a policy that is most effective in increasing access (reducing the participation cost) benefits the poor and talented agents primarily, while wealthy agents lose due to higher interest rates and wages. By contrast, policies that target financial depth (relaxing the borrowing constraint) benefit wealthy and talented agents but can impose losses on wealthy but less-talented agents.

Finally, a GDP decomposition shows that relaxing the credit participation cost increases GDP mainly through the extensive margin by enabling more entrepreneurs to obtain credit from banks. By contrast, relaxing the borrowing constraint or reducing the intermediation cost raises GDP mostly through the intensive margin by allowing entrepreneurs who are already in the credit regime to expand their businesses. Our TFP decomposition shows that there are large losses in TFP in the savings regime as talented entrepreneurs leave the savings regime when financial constraints are relaxed. More importantly, a large proportion of the increase in TFP generated by financial inclusion is due to a between-regime shifting effect, namely, talented but relatively poor entrepreneurs move from the savings to the credit regime and expand their businesses.

The remainder of the paper is organized as follows. The next section provides a brief overview of the related literature. Section 3 sets out the structure of the model. Section 4 highlights the differential impacts of relaxing different financial constraints and their interactions. Section 5 presents the data and the model calibration. Section 6 discusses the quantitative results. Finally, Section 7 provides concluding remarks.

2 Literature Review

A growing theoretical literature has emphasized the aggregate and distributional impacts of financial intermediation in models of occupational choice and financial frictions. Banerjee and Newman (1993) develop a framework with occupation choiceto capture the process of economic development; Lloyd-Ellis and Bernhardt (2000) extend the model to explain income inequality and the existence of a Kuznets curve. Cagetti and Nardi (2006) build on the framework to show that the introduction of a bequest motive generates lifetime savings profiles more consistent with data. In these studies, improved financial intermediation leads to greater entry into entrepreneurship, higher productivity and investment, and a general equilibrium effect that increases wages. Moreover, the models suggest that the distribution of wealth or the joint distribution of wealth and productivity is critical.

A related literature has found sizable impacts of improved financial intermediation on aggregate productivity and income (Gine and Townsend, 2004; Jeong and Townsend, 2007, 2008; Amaral and Quintin, 2010; Buera, Kaboski and Shin, 2011; Greenwood, Sanchez and Wang, 2013). Buera, Kaboski and Shin (2011) incorporate forward-looking agents in an occupational choice framework, and show that financial frictions account for a substantial part of the observed cross-country differences in output per worker and aggregate TFP. Moreover, Buera, Kaboski and Shin (2012) focus on the general equilibrium effects of micro finance. They find that the impact of scaling-up micro finance on per-capita income is small, because of the ensuing redistribution of income from high-savers to low-savers, but the vast majority of the population benefits from higher wages. Moll (2014) shows that the impact of financial frictions on GDP and TFP depends on the persistence of idiosyncratic shocks, and that the short-run effects of financial frictions tend to be larger than their long-run impacts.

Our model builds on this occupational choice framework, but with novel features. We focus on several dimensions of financial inclusion within an economy. Although these dimensions have typically been considered separately in the previous literature, our paper provides a unified framework for examining them individually as well as jointly. Our model features three types of financial frictions: fixed costs of credit entry, limited commitment, and asymmetric information. Unlike previous studies, our model allows us to also uncover how different frictions interact with each other. In this sense, our paper is related to studies in which multiple financial frictions co-exist and are compared. Clementi and Hopenhayn (2006) and Albuquerque and Hopenhayn (2004) argue that moral hazard and limited commitment have different implications for firm dynamics. Abraham and Pavoni (2005) and Doepke and Townsend (2006) discuss how consumption allocations differ under moral hazard with and without hidden savings versus full information. Martin and Taddei (2013) study the implications of adverse selection on macroeconomic aggregates and contrast them with those under limited commitment. Karaivanov and Townsend (2014) estimate the financial/information regime in place for households (including those running businesses) in Thailand and find that a moral hazard constrained financial regime fits the data best in urban areas, while a more limited savings regime is more applicable for rural areas. Similarly, Paulson, Townsend and Karaivanov (2006) argue that moral hazard best fits the data in the more urban Central region of Thailand but not in the more rural Northeast. Kinnan (2014) uses a different metric based on the first-order conditions characterizing optimal insurance under moral hazard, limited commitment, and hidden income to distinguish between these regimes in Thai data. Finally, Moll, Townsend and Zhorin (2014) use a general equilibrium framework that encompasses different types of frictions, and examine the equilibrium interactions among various frictions. Our paper is related to these studies, but we emphasize the rich interactions among financial constraints, which in partial equilibrium can be complements on the intensive margin and substitutes on the extensive margin. Our paper also constitutes a normative policy analysis. By developing a quantitative macroeconomic framework and disciplining it with micro data, we shed light on a number of policy issues. For instance, what financial frictions are most relevant for the economy's GDP and income inequality? And what is the impact of alleviating these financial frictions individually or jointly?

Our paper is also related to a large empirical literature on the real effects of credit. The view that financial inclusion spurs economic growth is supported by empirical evidence (King and Levine, 1993; Levine, 2005). Regression-based analyses at the aggregate level reveal a strong correlation between broad measures of financial depth (such as M2 or credit to GDP) and economic growth. For firms, access to finance is positively associated with innovation, job creation, and growth (Beck, Demirg-Kunt and Maksimovic, 2005; Ayyagari, Demirgc-Kunt and Maksimovic, 2008). There is also evidence that aggregate financial depth is positively associated with poverty reduction and income inequality (Beck, Demirg-Kunt and Levine, 2007; Clarke, Xu and fu Zou, 2006). Cross-sectional regression analysis, however, can be problematic as causality cannot easily be established, causal mechanisms are difficult to pin down, and policy evaluation is more challenging. Moreover, the implicit assumptions of stationarity and linearity in regression analysis could be incorrect, even after taking logs and including lags, if these variables lie on complex transitional growth paths (Townsend and Ueda, 2006). The advantage of using a structural framework such as ours lies in capturing salient features of the economy and the pertinent financial sector frictions.

Our paper is also broadly related to the literature on misallocation (Hsieh and Klenow, 2009; Caselli and Gennaioli, 2013; Midrigan and Xu, 2014; Moll, 2014) and inequality (Davies, 1982; Huggett, 1996; Aghion and Bolton, 1997; Castaneda, Diaz-Gimenez and Rios-Rull, 2003; Nardi, 2004). Our contribution is to show that policy options that target different financial sector frictions have different impacts on resource allocation and inequality. More importantly, even for the same policy, the impacts on inequality can differ due to country-specific characteristics.

3 The Model

The economy is populated by a continuum of agents of measure one. Agents are heterogeneous in terms of initial wealth b and talent z.

Agents live for two periods. In the first period, agents make credit participation, occupational choice, and investment decisions, taking the optimal consumption and bequest decisions made in the second period as given. In the second period, agents realize income as wages or business profits, depending on their occupations, and make consumption and bequest decisions to maximize utility. Each agent has an offspring, whose wealth is equal to the bequest, and talent is drawn from a stochastic process.[2] The time subscript t is omitted unless necessary.

3.1 Agents

Agents generate utility only in the second period through consumption and a bequest to their offspring. The utility function is Cobb-Douglas, given by

$$u(c, b') = c^{1-\omega} b'^{\omega}, \tag{3.1}$$

where c is consumption, and b' is bequest. The bequest motive transfers wealth across periods, which endogenously determines the economy's wealth distribution. The assumption that utility is generated by bequest rather than the offspring's utility simplifies the analysis and captures the idea of a tradition for bequest giving following Andreoni (1989).[3]

In the second period, agents maximize (3.1) by choosing c and b' subject to the budget constraint $c + b' = W$, where W denotes the second-period wealth, and it depends on the initial wealth and the realized first-period income.

The Cobb-Douglas form implies that the optimal bequest rate is ω.[4] Hence, the utility function $u(c, b')$ is a linear function of the end-of-period wealth (W), i.e., agents are risk neutral. This implies that maximizing expected utility is equivalent to maximizing expected second-period wealth. Therefore, in the first period, agents make credit participation decisions, occupational choices, and investment decisions to maximize expected income.

In the first period, agents need to make an occupational choice between being workers or entrepreneurs.[5] Each worker supplies one unit of labor, and the income realized in the first period

[2] The successor of an agent can be interpreted as the reincarnation of the original agent with potentially new talent.

[3] This is equivalent to assuming a myopic savings rate for the same agent. In Appendix B, we consider robustness checks and explore the implications of myopic savings rate by contrasting the simulation results in the baseline model with the results obtained from a model with forward-looking agents.

[4] The value of ω affects the amount of wealth transferred from the current period to the next period. Therefore, ceteris paribus a higher ω implies that the economy would have a higher level of wealth.

[5] In our framework, farmers can be considered as entrepreneurs, who operate their own farming businesses.

is equal to the equilibrium wage, w. Entrepreneurs employ capital and labor, and obtain income through business profits.

Talent is drawn from a Pareto distribution $\mu(z)$ with a tail parameter θ. The offspring inherits the talent of her parents (or former self) with probability γ, otherwise, a new talent is drawn from $\mu(z)$.[6]

Entrepreneurs have access to a production technology, the productivity of which depends on agents' talent. The production function is given by

$$f(k,l) = z(k^{\alpha}l^{1-\alpha})^{1-\nu}, \tag{3.2}$$

where $1-\nu$ is the Lucas span-of-control parameter, representing the share of output accruing to the variable factors. Out of this, a fraction α goes to capital, and $1-\alpha$ goes to labor. Production exhibits diminishing returns to scale, with $\nu > 0$. Capital depreciates by δ after use.

Production fails with probability p, in which case output is zero and agents are able to recover only a fraction $\eta < 1$ of installed capital, net of depreciation in the second period. To simplify the calculation, we assume workers get paid only when production is successful. Therefore, each worker earns a wage with probability $1-p$.

All agents can make a deposit in banks so as to transfer income and initial wealth across periods for consumption and bequest. However, following Greenwood and Jovanovic (1990) and Townsend and Ueda (2006), agents need to pay a fixed credit participation cost ψ to obtain a borrowing contract from banks. We assume that an agent lives in a "credit regime", if the agent pays the cost ψ and can borrow; that an agent lives in a "savings regime", if the agent does not pay ψ and can thereby only save. This cost can be considered as a contractual fee or a bargaining cost with banks. Intuitively, since workers do not invest, they never demand external credit. Entrepreneurs may want to borrow in order to expand their business scale and profits. In equilibrium, the fixed entry cost ψ is more likely to exclude poor entrepreneurs from financial markets, because this amounts to a larger fraction of their initial wealth. The next subsection illustrates the structure of the borrowing contract in detail.

Note that both the wage and the deposit rate are potentially time-varying and determined endogenously by the labor and capital market clearing conditions. Given the equilibrium wage w and deposit rate r^d, agents of type (b,z) make credit participation and occupational choice decisions to maximize expected income.

We solve the problem in two steps: first, agents choose their occupations conditional on the regime they are living in; second, agents choose the underlying regime by comparing the expected income that can be obtained in each regime. Next, we present the occupational choice problem in

[6] The shock to talent is interpreted as changes in market conditions that affect the profitability of individual skills as in Buera, Kaboski and Shin (2011).

the savings and credit regimes, respectively.

3.1.1 Savings Regime

Agents living in the savings regime cannot borrow from banks—they have to finance the production exclusively using their own resources.

In the first period, the goal of agents is to maximize expected income. Given a certain initial wealth, maximizing expected income is equivalent to maximizing expected end-of-period wealth, W. Let $\pi(b,z)$ be the expected end-of-period wealth function for entrepreneurs of type (b,z). Denoting variables in the savings regime with superscript S, one can write

$$W^S = \begin{cases} (1+r^d)b + (1-p)w & \text{for workers,} \\ \pi^S(b,z) & \text{for entrepreneurs,} \end{cases} \tag{3.3}$$

where workers are paid only if production is successful, with probability $(1-p)$. Since agents are risk neutral, they choose to be workers if $(1+r^d)b + (1-p)w > \pi^S(b,z)$, and entrepreneurs otherwise. Therefore, the end-of-period wealth can be simply written as $W^S = \max\{(1+r^d)b+(1-p)w, \pi^S(b,z)\}$.

The wealth function $\pi^S(b,z)$ for entrepreneurs is obtained from the following maximization problem

$$\begin{aligned} \pi^S(b,z) = \max_{k,l} \quad & (1-p)[z(k^\alpha l^{1-\alpha})^{1-\nu} - wl + (1-\delta)k] + p\eta(1-\delta)k + (1+r^d)(b-k), \\ \text{subject to} \quad & k \leq b. \end{aligned} \tag{3.4}$$

With probability $1-p$, production succeeds, and entrepreneurs get revenue, $z(k^\alpha l^{1-\alpha})^{1-\nu} - wl$, plus the undepreciated working capital, $(1-\delta)k$. With probability p, production fails, and entrepreneurs can only get a fraction η of the undepreciated working capital. The last term in the maximization problem accounts for the wealth that is not used in production, which earns the equilibrium interest rate r^d. The constraint reflects the fact that entrepreneurs need to finance capital through their own initial wealth. The optimal choice of capital and labor is characterized in Proposition 1.

Proposition 1. *In the savings regime, the optimal amount of capital invested by entrepreneurs of type (b,z) is given by*

$$\begin{aligned} k^*(b,z) &= \min(b, \tilde{k}^S(z)), \\ l^*(b,z) &= [\frac{z(1-\alpha)(1-\nu)}{w}]^{\frac{1}{\alpha(1-\nu)+\nu}} k^*(b,z)^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}, \end{aligned}$$

where $\tilde{k}^S(z) = [\frac{\alpha w(1-p)}{(1-\alpha)(r^d+(1-p)\delta - p\eta(1-\delta)+p)}]^{\frac{\alpha(1-\nu)+\nu}{\nu}}(\frac{(1-\nu)(1-\alpha)z}{w})^{\frac{1}{\nu}}$ is the unconstrained level of capital (scale of business) in the savings regime.

Note that $\tilde{k}^S(z)$ is the desired amount of capital that entrepreneurs living in the savings regime would like to invest when facing no wealth constraints. The value of $\tilde{k}^S(z)$ is finite because production has diminishing returns to scale. For entrepreneurs whose wealth is lower than $\tilde{k}^S(z)$, capital investment is constrained by wealth, i.e., $k^*(b, z) = b$.

3.1.2 Credit Regime

By paying an up-front credit participation cost ψ, agents enter the credit regime and obtain access to external credit. As workers do not need credit, they never pay ψ. Therefore, we only consider the entrepreneurs' problem in the credit regime.

We assume that the banking sector is perfectly competitive, driving the profit of intermediation to zero. This assumption can be easily relaxed by adding a profit margin for intermediation to capture noncompetitive banking sectors in many developing countries. This serves to increase the lending rate facing entrepreneurs, but the model's quantitative predictions would not change much.

In order to borrow, agents need to sign a contract with banks. A financial contract is characterized by three variables, (Φ, Δ, Ω), where Φ is the amount of borrowing, Δ is the value of collateral, and Ω is the face value of the contract. The face value Ω is the amount of money that needs to be repaid by the borrower if there is no default, which is determined by banks' zero profit condition. For simplicity, we assume that collateral is interest bearing, that is, agents earn the deposit rate r^d on the value of collateral.

Although the financial contract does not specify the lending rate, we can define the implied interest rate in the following way

$$r^l = \frac{\Omega}{\Phi} - 1. \tag{3.5}$$

Note that r^l would be potentially different for different entrepreneurs, depending on the terms of the contract.

Similarly, the leverage ratio (the amount of loans relative to the size of collateral) is defined as

$$\tilde{\lambda} = \frac{\Phi}{\Delta}. \tag{3.6}$$

If production fails, entrepreneurs may not be able to repay the loan's face value Ω. If this happens, entrepreneurs default and banks seize the interest-bearing collateral, $(1 + r^d)\Delta$, and the recovered value of undepreciated working capital, $\eta(1 - \delta)k$. In equilibrium, since highly-leveraged entrepreneurs default in the case of a production failure, they are charged with a higher lending rate in the event of success (to compensate for losses in the event of failure).

Limited commitment In order to borrow, entrepreneurs need to post collateral at banks. Suppose that entrepreneurs can borrow Φ if amounts of collateral Δ is posted. Suppose further that contract enforcement is imperfect, therefore, entrepreneurs can immediately abscond with a fraction $1/\lambda$ of the rented capital. The only punishment is that they lose their collateral Δ. In equilibrium, entrepreneurs do not abscond only if $\Phi/\lambda < \Delta$.[7] Therefore, banks are only willing to lend $\lambda\Delta$ to entrepreneurs if Δ units of collateral are posted. This single parameter $\lambda \geq 1$ parsimoniously captures the degree of financial friction resulting from limited commitment. A special case of $\lambda = 1$ implies that entrepreneurs cannot borrow.

Asymmetric information There is asymmetric information between entrepreneurs and banks (i.e. whether the production of a particular entrepreneur fails or not is only known to the entrepreneur). Due to limited liability, entrepreneurs have a default option when production fails. This implies that they could repay less if a production failure is reported and the lie is not discovered by banks. Banks have a monitoring technology through which they get information on the success of production at a cost proportional to the scale of the production (denoted by χ). If entrepreneurs are caught cheating, banks can legally enforce the full repayment of the loan's face value. As banks make zero profit in equilibrium, the monitoring cost is borne by entrepreneurs when the financial contract is designed. In sum, all agents are truth-telling. However, this comes at a cost.

The banks' optimal verification strategy follows Townsend (1979), which occurs if and only if entrepreneurs cannot repay the face value of the loan. This happens when entrepreneurs are highly leveraged and also experience a production failure.[8] To be more specific, when production succeeds, entrepreneurs can repay the face value of the loan.[9] Therefore, there is no incentive for banks to monitor. However, if a production failure is reported, banks monitor only if the loan contract is highly leveraged. This is because a low-leveraged loan contract implies that entrepreneurs are not borrowing much. Therefore, the required repayment is small, and can be covered by the value of interest-bearing collateral, $(1 + r^d)\Delta$, plus the value of recovered working capital, $\eta(1 - \delta)k$, even

[7] See Banerjee and Newman (2003), Buera and Shin (2013), and Moll (2014) for a similar motivation of this type of constraint. The borrowing constraint is derived based on the assumption that entrepreneurs can immediately walk away with the rented capital. Another possibility is that entrepreneurs may want to put this capital into production and walk away after output is realized. In this case, the condition that regulates diversion is $\Phi(R^k - R^l) + \Delta R^l \geq \frac{\Phi}{\lambda} R^k$, where R^k and R^d are the average gross return on capital and the gross lending rate, respectively. The implied borrowing constraint by this condition is $\Phi/\lambda < \frac{\Delta}{\lambda - (\lambda-1)R^k/R^d}$, which is more relaxed than the one in the main text. In fact, the two are equivalent only when $R^k = R^d$, which is the capital return obtained by the least talented entrepreneur. Since it is realistic to believe that banks cannot observe entrepreneurs' talent (an assumption we make later when discussing the optimal contract), it is reasonable to assume that banks would impose the most stringent borrowing constraint. As a result, the borrowing constraint derived from ex-post diversions is consistent with the borrowing constraint specified in the main text.

[8] Implicit here is the assumption that entrepreneurs would not decline the repayment of the loan if they have sufficient funds because banks monitor and seize the face value of the loan when default happens.

[9] To see this, notice that entrepreneurs would borrow to produce only if they can make profits. Therefore, when production succeeds, gross output should be at least higher than the capital input. On the other hand, if entrepreneurs default, banks monitor output and seize the face value of the loan anyway. Thus, entrepreneurs have no incentive to default.

if production fails. In this case, entrepreneurs have no incentive to lie because regardless of the production outcome, they can and have to repay the face value of the loan. For the same reason, banks have no incentive to monitor.

On the other hand, if the loan contract is highly leveraged[10], and if production fails, the amount that entrepreneurs can repay is not sufficient to cover the face value of the loan. As a result, default happens. Finally, note that in this case entrepreneurs have an incentive to lie when production is successful because they know that with high leverage, they would repay less if a production failure is reported. Therefore, to motivate truth-telling, banks verify all highly-leveraged loan contracts if a production failure is reported. We formalize the optimal verification strategy in Proposition 2.

Proposition 2. *Banks' optimal verification strategy is pinned down ex-ante and determined by the contract* $(\Phi,\ \Delta,\ \Omega)$*, parameter* η *and* δ*, and the deposit rate* r^d*:*

i. For a low-leveraged loan, i.e. $\eta(1-\delta)\Phi + (1+r^d)\Delta \geq \Omega$*, no verification occurs.*

ii. For a highly-leveraged loan, i.e. $\eta(1-\delta)\Phi + (1+r^d)\Delta < \Omega$*, verification occurs iff production fails.*

In the credit regime, the end-of-period wealth is denoted by

$$W^C = \pi^C(b, z),$$

where the superscript C refers to the credit regime. Agents choose to pay the credit participation cost when $W^C > W^S$.

We assume that banks cannot observe entrepreneurs' type (b, z), and therefore have to provide a menu of contracts. Entrepreneurs choose their optimal contracts from the menu. Notice that the schedule of contracts is designed to be incentive compatible, namely, entrepreneurs of type (b, z) would have no incentive to imitate type (b', z') and choose the optimal contract of other entrepreneurs. Moreover, all loan contracts make zero profit given that financial intermediation is perfectly competitive. Below, we first elaborate the optimal contract for entrepreneurs of type (b, z). We then discuss why the contract is incentive compatible.

To solve the optimal loan contract (Φ, Δ, Ω) for entrepreneurs of type (b, z), we use the following steps:

First, since collateral is interest-bearing, entrepreneurs are willing to post all of their wealth net of credit participation cost, $b - \psi$, as collateral instead of depositing a fraction of it in a savings account. Hence, the collateral term, $\Delta = b - \psi$, belongs to the set of optimal loan contracts.[11]

[10] The threshold between low and high leverage ratios is derived by considering whether the value of interest-bearing collateral plus the recovered working capital is sufficient to repay the face value of the loan. In particular, as we discuss later, the loan contract is highly leveraged if $\eta(1-\delta)\Phi + (1+r^d)\Delta < \Omega$.

[11] Note that there might exist multiple optimal contracts for wealthy entrepreneurs since they do not demand much credit. But all these contracts would result in an identical net outcome for both entrepreneurs and banks. The optimal contract we consider here is the one with the lowest leverage ratio, i.e., all wealth b is posted as collateral.

Second, entrepreneurs borrow to increase production scale and make higher profits. Therefore, there is no reason to borrow more funds from banks and not use them in production, since this would only increase the leverage ratio, which, in turn, potentially increases the cost of capital. Hence, the amount of loan Φ is equal to the amount of capital $k(b, z)$, if the loan contract is optimal.

The above arguments suggest that the optimal loan contract chosen by entrepreneurs of type (b, z) should be of the form $(k(b, z), b - \psi, \Omega)$. Hence, Ω remains the only element to be determined.

The face value of the loan Ω in the optimal contract is set such that banks make zero profit knowing that only entrepreneurs of type (b, z) will choose it. From banks' perspective, the expected payoff of this loan contract is $(1-p)\Omega + p\min(\Omega, \eta(1-\delta)k + (1+r^d)(b-\psi))$. The first term refers to the payoff when production succeeds, which happens with probability $1-p$. In this case, banks receive the full face value of the loan, Ω. The second term refers to the payoff when production fails. When production fails, before repaying debt, entrepreneurs' "net value" is equal to the recovered working capital, $\eta(1-\delta)k$, plus the after-interest value of collateral, $(1+r^d)(b-\psi)$. Banks receive the full face value of the loan, Ω, if entrepreneurs' "net value" is sufficient to repay it. Otherwise, banks only receive the "net value" due to limited liability, and entrepreneurs would end up with nothing. In sum, when production fails, banks receive either Ω or $\eta(1-\delta)k + (1+r^d)(b-\psi)$, whichever is smaller.

On the other hand, the cost of creating the loan contract is equal to the after-interest value of the loan, $(1+r^d)k$, plus the expected cost of monitoring. Note that monitoring occurs only if entrepreneurs cannot repay the loan, namely, when production fails and the net value, $\eta(1-\delta)k + (1+r^d)(b-\psi)$, is smaller than the loan's face value, Ω. In this case, a monitoring cost, χk, is incurred. Therefore, the expected cost of monitoring is equal to the monitoring cost, χk, multiplied by the monitoring rate. The monitoring rate is equal to the production failure rate, p, when entrepreneurs are highly leveraged, i.e. $\eta(1-\delta)k + (1+r^d)(b-\psi) < \Omega$, and zero otherwise. Thus the expected cost of monitoring can be expressed as $p\chi k \cdot \mathbb{1}_{\{\eta(1-\delta)k+(1+r^d)(b-\psi)<\Omega\}}$, where $\mathbb{1}_{\{\eta(1-\delta)k+(1+r^d)(b-\psi)<\Omega\}}$ is an indicator function, which equals to 1 if $\eta(1-\delta)k + (1+r^d)(b-\psi) < \Omega$ and 0 otherwise. Hence, the cost of creating the loan contract is $(1+r^d)k + p\chi k \cdot \mathbb{1}_{\{\eta(1-\delta)k+(1+r^d)(b-\psi)<\Omega\}}$.

The zero profit function is obtained when the expected payoff of the loan is equal to its cost:

$$(1-p)\Omega + p\min(\Omega, \eta(1-\delta)k + (1+r^d)(b-\psi)) = (1+r^d)k + p\chi k \cdot \mathbb{1}_{\{\eta(1-\delta)k+(1+r^d)(b-\psi)<\Omega\}}. \tag{3.7}$$

Equation (3.7) pins down Ω, and implies that in the optimal contract we consider, Ω is a function of k and b only. The optimal contract chosen by entrepreneurs of type (b, z) can be written as $(k^*(b, z), b-\psi, \Omega(k^*(b, z), b))$, where $k^*(b, z)$ is the optimal amount of capital invested in production, and $\Omega(k^*(b, z), b)$ is determined by equation (3.7). This implies that to exactly characterize the optimal contract as a function of initial variables b and z, we only need to know $k^*(b, z)$, which solves the following problem:

$$\begin{aligned} \pi^C(b,z) = \max_{k,l} \quad & (1-p)[z(k^\alpha l^{1-\alpha})^{1-\nu} - wl + (1-\delta)k - \Omega + (1+r^d)(b-\psi) \\ & +p\max(0, \eta(1-\delta)k + (1+r^d)(b-\psi) - \Omega), \\ \text{subject to} \quad & k \ \leq \ \lambda(b-\psi), \end{aligned} \tag{3.8}$$

where the term Ω in problem (3.8) is the solution to banks' zero profit condition (3.7). The solution to (3.7) and (3.8) determines the optimal capital k as a function of b and z, and pins down the optimal contract.

In (3.8), the first term refers to the end-of-period wealth when production succeeds. The second term refers to the case of production failure. Entrepreneurs have something left only if $\eta(1-\delta)k + (1+r^d)(b-\psi) > \Omega$, that is when the recovered undepreciated working capital plus the after-interest value of collateral is sufficient to repay the loan. Otherwise, entrepreneurs end up with zero end-of-period wealth.

Below we restrict ourselves to the case where default occurs, with the endogenously determined interest rate satisfying, $r^d > \dfrac{\eta(1-\delta)\lambda}{\lambda-1} - 1$.[12] Note that this condition is satisfied for all the six countries in our quantitative analysis. We first illustrate the default boundary (Lemma 1) and the associated cost of capital for different cases (Lemma 2), and then we characterize the optimal amount of capital in Proposition 3.

Lemma 1. *In the credit regime, default occurs for highly-leveraged entrepreneurs. In particular, there is a default boundary,* $\overline{\lambda} = \dfrac{1+r^d}{1+r^d-\eta(1-\delta)}$*, depending on parameters* η *and* δ *and the endogenous deposit rate* r^d*. For an entrepreneur who operates a business with leverage ratio* $\tilde{\lambda}$*:*

i. *If* $\tilde{\lambda} \leq \overline{\lambda}$ *(low-leverage region), default never occurs, and the implied lending rate is* $r^l = r^d$.

ii. *If* $\tilde{\lambda} > \overline{\lambda}$ *(high-leverage region), default occurs when production fails, and the implied lending rate is increasing in* $\tilde{\lambda}$*, i.e.* $r^l = \dfrac{1+r^d+p\chi-p\eta(1-\delta)-p(1+r^d)/\tilde{\lambda}}{1-p} - 1$.

Lemma 1 states that default happens only for highly-leveraged entrepreneurs whose production fails. Moreover, for entrepreneurs with no default risk (i.e., $\tilde{\lambda} \leq \overline{\lambda}$), banks can always get repaid the face value of the loan, and the implied lending rate r^l is equal to the deposit rate r^d. For entrepreneurs facing a risk of default (i.e., $\tilde{\lambda} > \overline{\lambda}$), the implied lending rate is increasing in the leverage ratio to compensate for losses from default. In general, for highly-leveraged entrepreneurs, the lending rate includes a risk premium which depends on the leverage ratio and the fixed intermediation cost from bank monitoring.

[12] If $r^d \leq \dfrac{\eta(1-\delta)\lambda}{\lambda-1} - 1$, there is no default in the economy. This is because in our model, whether an entrepreneur defaults or not depends on the leverage ratio. As shown in Lemma 1, only entrepreneurs whose leverage ratios are larger than $\overline{\lambda}$ default when production fails. Notice that $\overline{\lambda}$ is decreasing in r^d. Therefore, $\overline{\lambda}$ could be higher than λ (the highest possible leverage ratio imposed by limited commitment) for small r^d. In this case, even entrepreneurs with fully leveraged loans do not default.

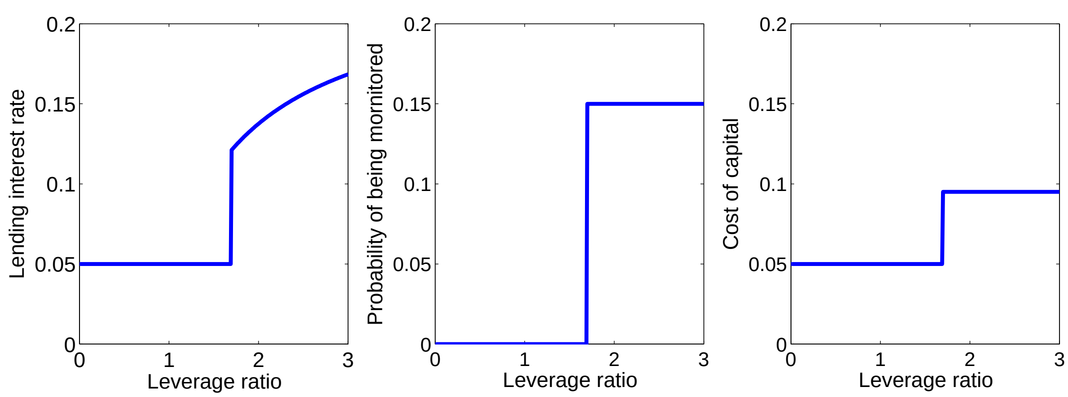
Note that the implied lending rate is not equal to the cost of capital facing entrepreneurs. The lending rate should be considered as the interest rate entrepreneurs need to pay when production is successful. But if production fails, entrepreneurs have the option to default and pay less. The cost of capital includes this default option. Therefore, it is a weighted average of the lending rate and the repayment rate during default. This is characterized in Lemma 2.

Lemma 2. *In the credit regime, for an entrepreneur who operates a business with leverage ratio* $\tilde{\lambda}$*:*

i. If $\tilde{\lambda} \leq \overline{\lambda}$*, the cost of capital is* r^d*.*

ii. If $\tilde{\lambda} > \overline{\lambda}$*, the cost of capital is* $r^d + p\chi$*.*

In Figure 3, we show how the lending rate, the probability of being monitored, and the cost of capital change when the leverage ratio varies. As noted in Proposition 2, only highly-leveraged entrepreneurs are monitored. In particular, there is a default boundary, $\overline{\lambda} = 1.69$, below which the probability of being monitored is zero, and thus both the lending rate and the cost of capital are equal to the deposit rate. If entrepreneurs increase leverage beyond this boundary, they cannot repay the face value of the loan when production fails. Therefore, the probability of being monitored is exactly equal to the production failure rate, p. Since banks are making zero profit, the monitoring cost is completely borne by entrepreneurs, generating a higher cost of capital. Note that the cost of capital in this case is $r^d + p\chi$, which is constant regardless of the leverage ratio (see Lemma 2). This is due to our assumption that the monitoring cost is proportional to the scale of production but not the value of the loan. Moreover, the implied lending rate characterized in Lemma 1 is strictly increasing in the leverage ratio when the leverage ratio is higher than the default boundary. This is because banks have to be repaid more (as reflected by a higher face value Ω) when production succeeds to compensate for larger losses during production failure arising from higher leverage.



Note: The left panel plots the implied lending rate against the leverage ratio; the middle panel plots the monitoring frequency against the leverage ratio; the right panel plots the implied cost of capital against the leverage ratio. All panels are plotted using the following parameter values: $r^d = 0.05$, $\eta = 0.35$, $\delta = 0.06$, $p = 0.15$, $\chi = 0.3$.

Figure 3: The lending rate, the monitoring frequency, and the cost of capital for different leverage ratios.

Next we characterize the optimal amount of capital invested by entrepreneurs of type (b, z).

Proposition 3. *In the credit regime, for entrepreneurs of type (b, z), denote the optimal leverage ratio by $\lambda^*(b, z)$ and optimal capital by $k^*(b, z)$. There is a threshold level of wealth $\bar{b}(z)$, such that:*

i. *If wealth b is between the participation cost and the threshold level, $\psi \leq b < \bar{b}(z)$, the optimal leverage ratio lies between the default boundary and the inverse of the absconding rate,*

$$\begin{aligned} \bar{\lambda} &< \lambda^*(b, z) \leq \lambda, \\ k^*(b, z) &= \min(\lambda(b - \psi), \tilde{k}^h(z)), \end{aligned}$$

where $\tilde{k}^h(z)$ is defined in (iii) below.

ii. *If wealth b is above the threshold level, $b \geq \bar{b}(z)$, the optimal leverage ratio is below the default boundary,*

$$\begin{aligned} \lambda^*(b, z) &\leq \bar{\lambda}, \\ k^*(b, z) &= \min(\bar{\lambda}(b - \psi), \tilde{k}^l(z)), \end{aligned}$$

where $\tilde{k}^l(z)$ is defined in (iii) below.

iii. *$\tilde{k}^h(z)$ is the unconstrained level of capital in the high-leverage region,*

$$\tilde{k}^h(z) = [\frac{(1-p)\alpha w}{(r^d + p\chi + (1-p)\delta - p\eta(1-\delta) + p)(1-\alpha)}]^{\frac{\alpha(1-\nu)+\nu}{\nu}} (\frac{(1-\nu)(1-\alpha)z}{w})^{\frac{1}{\nu}}.$$

$\tilde{k}^l(z)$ is the unconstrained level of capital in the low-leverage region,

$$\tilde{k}^l(z) = [\frac{(1-p)\alpha w}{(r^d + (1-p)\delta - p\eta(1-\delta) + p)(1-\alpha)}]^{\frac{\alpha(1-\nu)+\nu}{\nu}} (\frac{(1-\nu)(1-\alpha)z}{w})^{\frac{1}{\nu}}.$$

Note that $k^h(z) < k^l(z)$ for all z. This is because in the high-leverage region, banks monitor when production fails, which increases the cost of capital. When entrepreneurs are constrained by wealth, increasing the leverage ratio can generate higher revenue, but this may also push them into the "default region", increasing their cost of capital. Entrepreneurs want to maximize profits, but are always facing this trade-off when making investment decisions. For entrepreneurs with low wealth, the marginal return on capital is high. The extra revenue generated by increasing leverage beyond $\bar{\lambda}$ outweighs the increase in the cost of capital, hence they choose higher leverage ($\tilde{\lambda} > \bar{\lambda}$). By contrast, for relatively wealthy entrepreneurs, the marginal return on capital is low. As a result, they choose to borrow less and stay in the low-leverage region to avoid paying the monitoring cost.

Our model features both limited commitment and asymmetric information. In a model with only limited commitment, the supply of credit is rationed exogenously by the parameter λ. When asymmetric information is introduced, since monitoring is costly, in equilibrium there are some entrepreneurs who restrain themselves from borrowing more. For these entrepreneurs, the borrowing constraint imposed by limited commitment is not binding. In fact, they are restricting themselves from using up the credit line precisely because obtaining more credit brings them into the high-leverage region and increases their cost of capital. In this sense, credit rationing is endogenously imposed by entrepreneurs themselves.

Intuitively, the return on production is higher for talented entrepreneurs, which induces them to leverage more. This leads to Proposition 4.

Proposition 4. *The threshold level of wealth $\bar{b}(z)$ is increasing in z.*

Finally, all contracts offered by banks are incentive compatible, although talent is not observable. This implies that entrepreneurs with low talent have no incentive to pretend to be highly talented and ask for a different contract, or vice versa. To see this, divide both sides of equation (3.7) by k,

$$(1-p)\frac{\Omega}{k} + p\min(\frac{\Omega}{k}, \eta(1-\delta) + (1+r^d)\frac{b-\psi}{k}) = (1+r^d) + p\chi \cdot 1_{\{\eta(1-\delta)+(1+r^d)\frac{b-\psi}{k} < \frac{\Omega}{k}\}}. \tag{3.9}$$

Equation (3.9) suggests that the implied gross lending rate, $\frac{\Omega}{k}$, depends only on the inverse of the leverage ratio $\frac{b-\psi}{k}$[13], but not directly on entrepreneurs' talent. That is, capital k and talent z enter equation (3.9) only through the leverage ratio, which is observable. Therefore, for all entrepreneurs, given the amount of capital they want to invest (or demand for credit) and the amount of wealth they own (or collateral value), it is impossible to receive a lower interest rate from banks by cheating on talent. This result is obtained because it is assumed that the recovered value of undepreciated working capital does not depend on entrepreneurs' talent.

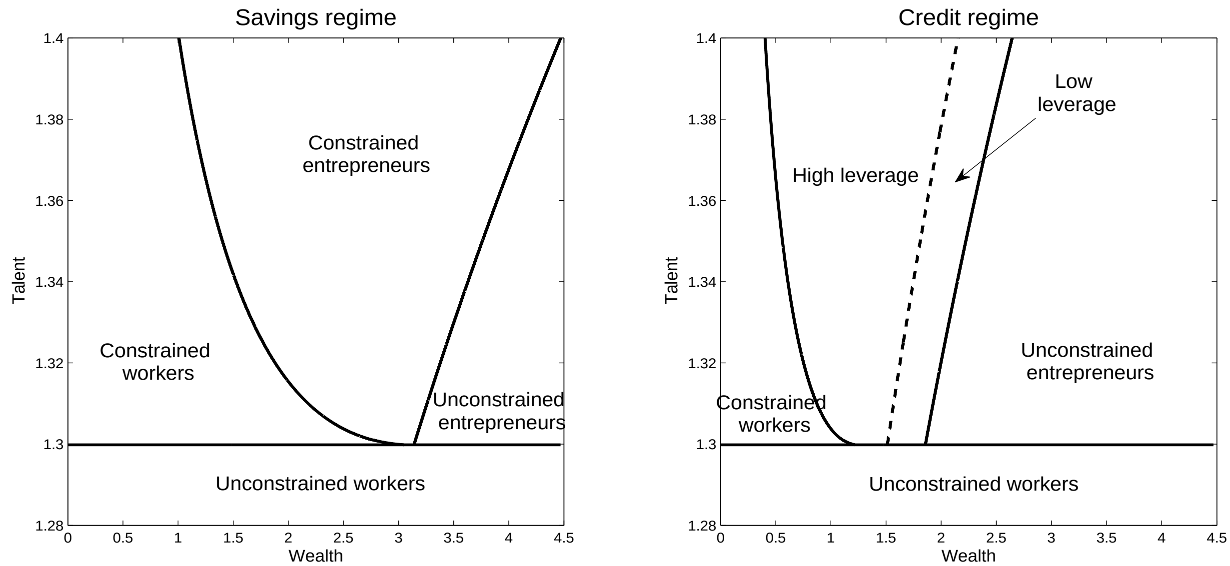
3.1.3 Occupational Choice

The occupation map is plotted based on the choice of occupation for agents with different talent z and wealth b, and whether this choice is constrained by wealth. We identify four categories of agents in the savings regime, separated by the solid lines in the left panel of Figure 4: unconstrained workers, constrained workers, constrained entrepreneurs, and unconstrained entrepreneurs.

As shown in the figure, there is a certain threshold level of talent (1.3) below which agents always find working for a wage better than operating a business. These agents are identified as unconstrained workers, suggesting that their talent is so low that they never find it optimal to become entrepreneurs. Above this talent level, the figure is further segmented into three regions. In the left region, agents are talented, but do not have sufficient wealth, so they cannot operate businesses at a profitable scale. Hence, they choose to be workers. These are constrained workers.

[13] According to (3.6), the inverse of leverage ratio is defined as $\frac{\Delta}{\Phi}$. In the optimal contract illustrated above, $\Delta = b - \psi$, and $\Phi = k$.

The middle region represents agents with sufficient wealth to operate profitable businesses but scale is still constrained by wealth ($k^*(b,z) < k^s(z)$). These agents are constrained entrepreneurs. Agents in the right region of the figure choose to be entrepreneurs, operating businesses at the unconstrained scale ($k^*(b,z) = k^s(z)$), with the marginal return on capital equal to the deposit rate. Thus, they are identified as unconstrained entrepreneurs.



Note: The left panel plots the occupation choice map in the savings regime, which is partitioned into four regions depending on talent and wealth: unconstrained/constrained workers, unconstrained/constrained entrepreneurs. The right panel plots the occupation choice map in the credit regime. The region of constrained entrepreneurs is further partitioned into entrepreneurs with high leverage ratios and entrepreneurs with low leverage ratios. All panels are plotted using the following parameter values: $r^d = 0.05$, $w = 0.6$, $\eta = 0.35$, $\delta = 0.06$, $\nu = 0.21$, $p = 0.15$, $\alpha = 0.33$, $\lambda = 2.5$, $\psi = 0$, $\chi = 0$.

Figure 4: The occupation choice map in the two regimes.

When agents obtain external credit and enter the credit regime, the occupation map changes to the one in the right panel of Figure 4. The occupation map for the credit regime is plotted with the same parameter values, and under the assumption that there is no credit participation cost, $\psi = 0$, or monitoring cost, $\chi = 0$.[14] This is to highlight the effect of external credit. Clearly, the area of constrained workers shrinks and that of unconstrained entrepreneurs increases. This implies that agents are more likely to become entrepreneurs and operate their businesses at a larger scale once credit is obtained from banks. Note that the region of constrained entrepreneurs is further partitioned by the dotted line into two sub-regions depending on their leverage ratios. Agents in the low-leverage region are not borrowing much in the sense that the face value of the loan can be repaid even if production fails. Thus, banks do not monitor them, and the lending rate is equal to the deposit rate, as shown in Figure 3. By contrast, agents in the high-leverage region default

[14] Note that we also use the same wage and interest rate while plotting the occupation choice map for the credit regime. This is to highlight the partial equilibrium result of moving an agent from the savings regime to the credit regime. When financial inclusion allows more agents to get credit, the wage and interest rate would also change in general equilibrium.

when production fails, in which case banks monitor and seize the recovered undepreciated working capital and after-interest collateral. In accordance with Proposition 3, the high-leverage region is to the left of the low-leverage region, implying that entrepreneurs prefer to leverage more when wealth is low given the high marginal return on capital.

The policy options we consider in Section 6, move the lines in the occupation map (Figure 4), and alter the relative income received by different agents. This kind of micro-level adjustment for each agent impacts the aggregate economy and generates a movement in GDP and income inequality.

3.2 Competitive Equilibrium

Given an initial joint probability density distribution of wealth and talent $h_0(b,z)$, a competitive equilibrium consists of allocations $\{c_t(b,z), k_t(b,z), l_t(b,z)\}_{t=0}^{\infty}$, sequences of joint distributions of wealth and talent $\{h_t(b,z)\}_{t=1}^{\infty}$ and prices $\{r^d(t), w(t)\}_t$, such that:

(1). Agents of type (b,z) optimally choose the underlying regime, occupations, consumption $c_t(b,z)$, capital $k_t(b,z)$, and labor $l_t(b,z)$ to maximize utility at $t \geq 0$.

(2). The capital market clears at all $t \geq 0$,

$$\iint_{(b,z)\in E(t)} k_t(b,z)h_t(b,z)dbdz = \iint_{(b,z)} bh_t(b,z)dbdz - \psi \iint_{(b,z)\in Fin(t)} h_t(b,z)dbdz,$$

where $E(t)$ is the set of agents, who choose to be entrepreneurs at time t; $Fin(t)$ is the set of agents, who are in the credit regime.

(3). The labor market clears at all $t \geq 0$,

$$\iint_{(b,z)\in E(t)} l_t(b,z)h_t(b,z)dbdz = \iint_{(b,z)\notin E(t)} h_t(b,z)dbdz.$$

(4). $\{h_t(b,z)\}_{t=1}^{\infty}$ evolves according to the equilibrium mapping.

$$h_{t+1}(\bar{b},\bar{z})db = \gamma\mu(\bar{z}) \int_z \mathbb{1}_{\{b'=\bar{b}\}} h_t(b,z)dbdz + (1-\gamma) \int_b \mathbb{1}_{\{b'=\bar{b}\}} h_t(b,\bar{z})db,$$

where b' is the bequest for agents of type (b,z), and $\mathbb{1}_{\{b'=\bar{b}\}}$ is an indicator function which equals 1 if $b' = \bar{b}$, and equals 0 otherwise.

The steady state of the economy is defined as the invariant joint distribution of wealth and talent $h(b,z)$,

$$h(b,z) = \lim_{t\to\infty} h_t(b,z).$$

4 Distinguishing the Impact of Financial Constraints

In this section, we explore the impact of different financial constraints on GDP when these constraints are relaxed separately and in combination. We first provide a partial equilibrium analysis focusing on individuals' output and credit access conditions. This enables us to uncover the underlying mechanisms of the model and distinguish the impact of different financial constraints. We then decompose GDP and TFP to shed light on the macroeconomic effects of financial inclusion.

Financial inclusion is reflected by three parameters in our model. The credit participation cost ψ directly measures the difficulty of obtaining credit. A decrease in its value therefore reflects greater financial access. The borrowing constraint parameter λ coincides directly with the maximum leverage ratio, an increase in which reflects lower collateral requirements. Finally, a decrease in the monitoring cost χ indicates an increase in the efficiency of financial intermediation.

Because financial inclusion is multidimensional, it is difficult to precisely identify these three parameters from an empirical standpoint. However, one can find evidence of policies that address one dimension or the other. For example, Assuncao, Mityakov and Townsend (2012) and Alem and Townsend (2013) find that the distance to a bank branch matters for credit access, which suggests that policies that promote branch openings in rural unbanked locations could help reduce the credit participation cost ψ in our model.[15] Moreover, during the recent financial crisis, many countries widened the range of securities that could be accepted as collateral with the aim of boosting lending to firms and households. This reflects an increase in λ in our model. Finally, financial liberalization and the resultant competition between financial institutions could accelerate investment in computerization, thereby improving intermediation efficiency (as reflected by a decrease in χ in our model). For example, from 1985 to 1994, the Thai banking sector had become a more capital-intensive industry, substituting physical capital for labor. The average cost of raising funds decreased from 14.40% in 1985 to 5.61% in 1994 for large-sized banks (Okuda and Mieno, 1999).

We distinguish the effect of financial inclusion on the extensive margin and the intensive margin. On the one hand, relaxing financial constraints can increase GDP through the extensive margin by increasing the credit access ratio (i.e., moving entrepreneurs from the savings regime to the credit regime). On the other hand, relaxing financial constraints enable entrepreneurs in the credit regime to produce more output, which boosts GDP. This effect operates on the intensive margin.

[15] Many developing countries have conducted such kind of policies. For example, after a bank nationalization in 1969, the Indian government launched an ambitious social banking program which sought to improve the access of the rural poor to formal credit and savings opportunities (Burgess and Pande, 2005).

4.1 The Impact at the Individual Level

We focus on the partial equilibrium with interest rates and wages fixed. We consider constant returns to scale production function (i.e., $\nu = 0$) to simplify the algebra. All the results also hold in the general case with decreasing returns to scale production function, but at the expense of losing closed-form solutions.

When $\nu = 0$, the threshold level of wealth, $\overline{b}(z)$, which separates "high-leverage" and "low-leverage" regions in Proposition 3 only takes two values, 0 or ∞, as shown in Lemma 3.

Lemma 3. *In the credit regime with $\nu = 0$, for an entrepreneur of type (b, z), there exists a threshold of talent $\overline{z}$, such that:*

i. If $z > \overline{z}$, then $\overline{b}(z) = \infty$ and the optimal leverage ratio is $\lambda^*(b, z) = \lambda$.

ii. If $z \leq \overline{z}$, then $\overline{b}(z) = 0$ and the optimal leverage ratio is $\lambda^*(b, z) = \overline{\lambda}$.

iii. The threshold of talent is $\overline{z} = \frac{w}{1-\alpha}[(\frac{p\chi\lambda}{(\lambda - \overline{\lambda})} + (1+r^d) + p\eta(1-\delta) - (1-p)(1-\delta))\frac{1-\alpha}{\alpha w(1-p)}]^{\alpha}$.

In fact, Lemma 3 can be considered as a corollary of Proposition 4 for the case where $\nu = 0$. Talented entrepreneurs face a steeper profit function, therefore they would like to choose higher leverage ratios. When the production function exhibits constant returns to scale, we obtain a "bang-bang" solution for the wealth threshold $\overline{b}(z)$ since the unconstrained level of capital is infinite.

Denote $\underline{b}(\psi, \lambda, \chi; z)$ as the threshold of wealth above which entrepreneurs of type (b, z) choose to enter the credit regime. A lower $\underline{b}(\psi, \lambda, \chi; z)$ implies that, all else equal, entrepreneurs with talent z are more likely to enter the credit regime.

In the following theorem, we show that relaxing financial constraints (decreasing ψ, increasing λ, or decreasing χ) can have differential quantitative impacts on reducing agents' wealth thresholds $\underline{b}(\psi, \lambda, \chi; z)$, reflecting their effect on the extensive margin.

Theorem 1. ***The Impact of Financial Constraints on the Extensive Margin***
In the credit regime with fixed interest rates and wages, and when $\nu = 0$:

i. Relaxing each financial constraint improves credit access:

$$-\frac{\partial \underline{b}}{\partial \psi} \leq 0, \quad \frac{\partial \underline{b}}{\partial \lambda} \leq 0, \quad -\frac{\partial \underline{b}}{\partial \chi} \leq 0.$$ [16]

ii. Financial constraints are substitutes on the extensive margin:

$$-\frac{\partial^2 \underline{b}}{\partial \psi \partial \lambda} \geq 0, \quad -\frac{\partial^2 \underline{b}}{\partial \lambda \partial \chi} \geq 0, \quad \frac{\partial^2 \underline{b}}{\partial \chi \partial \psi} \geq 0. \tag{4.1}$$

[16] There is a negative sign in front of $\frac{\partial \underline{b}}{\partial \psi}$ and $\frac{\partial \underline{b}}{\partial \chi}$ since relaxing the two constraints implies reducing the credit participation cost and the intermediation cost.

Theorem 1.i indicates that relaxing any constraint can unambiguously reduce the wealth threshold $\underline{b}(\psi, \lambda, \chi; z)$, which facilitates credit access (as captured by the first derivatives). The exact quantitative impacts, however, depend on individual characteristics (b, z) and country-specific parameters $(p, \eta, \delta, \alpha)$ as presented in Appendix A.7.[17] Note that the underlying mechanisms for relaxing different constraints are not identical. Lowering the credit participation cost ψ induces entrepreneurs to enter the credit regime by decreasing the ex-ante cost of borrowing, while a lower intermediation cost χ reduces the ex-post cost of borrowing. Relaxing the borrowing constraint λ motivates entrepreneurs to obtain credit by increasing their profits in the credit regime.

Importantly, Theorem 1.ii indicates that the three financial constraints are pair-wise substitutes on the extensive margin. For example, a lower credit participation cost dampens the effect of relaxing the borrowing constraint or reducing the intermediation cost. This is because a lower credit participation cost results in a lower wealth threshold $\underline{b}(\psi, \lambda, \chi; z)$, thus relaxing the borrowing constraint and reducing the intermediation cost have less of an impact on further reducing this threshold. In other words, when the credit participation cost is low, the credit access ratio is already high, so with little room for increasing this ratio further through the other two channels. Essentially, the substitution effect arises due to the natural bound on the maximum credit access ratio, which is 100%.

Financial inclusion increases agents' well being not only through its impact on promoting credit access (the extensive margin), but also by increasing the net output of entrepreneurs living in the credit regime (the intensive margin).

We define entrepreneurs' net output as the expected output net of direct costs arising from financial frictions, if any.[18] Thus in the savings regime, the net output $y^S(\psi, \lambda, \chi; b, z)$ is equal to the difference between the end-of-period wealth and the beginning-of-period wealth plus the user cost of capital and the labor cost:

$$y^S(\psi, \lambda, \chi; b, z) = \pi^S(b, z) + (r^d + \delta)k^*(b, z) + (1 - p)wl^*(b, z) - (1 + r^d)b, \tag{4.2}$$

where $\pi^S(b, z)$, $k^*(b, z)$, and $l^*(b, z)$ are solutions to problem (3.4).

In the credit regime, the net output $y^C(\psi, \lambda, \chi; b, z)$ is

$$y^C(\psi, \lambda, \chi; b, z) = \pi^C(b, z) + (r^d + \delta)k^*(b, z) + (1 - p)wl^*(b, z) - (1 + r^d)b, \tag{4.3}$$

where $\pi^C(b, z)$, $k^*(b, z)$, and $l^*(b, z)$ are solutions to problem (3.8).

[17] For the case with $z \leq \underline{z}$, the impact of relaxing the borrowing constraint λ has no impact on the wealth threshold (i.e., $\dfrac{\partial \underline{b}}{\partial \lambda} = 0$). This is because entrepreneurs with $z \leq \underline{z}$ choose the leverage ratio $\overline{\lambda} = \dfrac{1 + r^d}{1 + r^d - \eta(1 - \delta)}$ (see Lemma 3). Therefore, parameter λ can only affect $\overline{\lambda}$ through its impact on the equilibrium interest rate r^d, which is ruled out in our partial equilibrium analysis. Similar arguments also apply to Theorem 2.

[18] The net output defined in this way enables us to calculate GDP as the sum of all entrepreneurs' net output.

The next theorem presents the impact of relaxing financial constraints on agents' net output in the credit regime.

Theorem 2. ***The Impact of Financial Constraints on the Intensive Margin***
In the credit regime with fixed interest rates and wages, and when $\nu = 0$*:*

i. Relaxing each financial constraint raises net output:

$$-\frac{\partial y^C}{\partial \psi} \geq 0, \quad \frac{\partial y^C}{\partial \lambda} \geq 0, \quad -\frac{\partial y^C}{\partial \chi} \geq 0.$$

ii. Financial constraints are complements on the intensive margin:

$$-\frac{\partial^2 y^C}{\partial \psi \partial \lambda} \geq 0, \quad -\frac{\partial^2 y^C}{\partial \lambda \partial \chi} \geq 0, \quad \frac{\partial^2 y^C}{\partial \chi \partial \psi} \geq 0. \tag{4.4}$$

Theorem 2.i indicates that relaxing any constraint increases entrepreneurs' net output in the credit regime. However, it is important to emphasize that the exact quantitative impacts are different (see Appendix A.8). Theorem 2.ii says that relaxing any two constraints has complementary effects in boosting output. For example, when the credit participation cost is lower, entrepreneurs are left with more wealth after entering the credit regime. Since both the amount of credit and the total intermediation cost are proportional to wealth, relaxing the borrowing constraint and reducing the intermediation cost increases business profits by more.

In sum, the above partial equilibrium discussions provide important policy implications. First, financial inclusion policies matter since the quantitative effects differ when different financial constraints are relaxed. In general, the quantitative effects depend on the joint distribution of agents' wealth and talent and country-specific characteristics. If policy makers are constrained in using a single policy, then choosing the right policy to address the "bottleneck" constraint is tempting. Second, different constraints are complements on the intensive margin and substitutes on the extensive margin. Therefore, an optimal combination of policies is necessary in order to boost GDP.

4.2 The Impact on the Aggregate Economy

We now discuss the general equilibrium impact of financial constraints on the aggregate economy.

4.2.1 GDP Decomposition

The economy's GDP can be written as the sum of net output produced by entrepreneurs in the savings regime and those in the credit regime,

$$\text{GDP} = \underbrace{\int_z \int_0^{\bar{b}(\psi,\lambda,\chi;z)} y^S(\psi,\lambda,\chi;b,z)h(b,z)dbdz}_{\text{Total output in savings regime}} + \underbrace{\int_z \int_{\bar{b}(\psi,\lambda,\chi;z)}^{\infty} y^C(\psi,\lambda,\chi;b,z)h(b,z)dbdz}_{\text{Total output in credit regime}}. \tag{4.5}$$

Note that entrepreneurs' output in the savings regime $y^S(\psi,\lambda,\chi;b,z)$ is indirectly affected by financial parameters (ψ,λ,χ) through changes in equilibrium wages and interest rates. When financial constraints are relaxed from (ψ,λ,χ) to (ψ',λ',χ') with $\psi' \leq \psi$, $\lambda' \geq \lambda$, and $\chi' \leq \chi$, the increase in GDP can be decomposed into three margins as follows:

$$\begin{aligned}
\Delta\text{GDP}_{(\psi,\lambda,\chi)\text{ to }(\psi',\lambda',\chi')} = & \underbrace{\int_z \int_{\bar{b}(\psi',\lambda',\chi';z)}^{\bar{b}(\psi,\lambda,\chi;z)} [y^C(\psi',\lambda',\chi';b,z) - y^S(\psi,\lambda,\chi;b,z)]h(b,z)dbdz}_{\text{Gains on the extensive margin}} \\
& + \underbrace{\int_z \int_{\bar{b}(\psi,\lambda,\chi;z)}^{\infty} [y^C(\psi',\lambda',\chi';b,z) - y^C(\psi,\lambda,\chi;b,z)]h(b,z)dbdz}_{\text{Gains on the intensive margin}} \\
& - \underbrace{\int_z \int_0^{\bar{b}(\psi',\lambda',\chi';z)} [y^S(\psi,\lambda,\chi;b,z) - y^S(\psi',\lambda',\chi';b,z)]h(b,z)dbdz}_{\text{General equilibrium effects in the savings regime}}.
\end{aligned} \tag{4.6}$$

First, more entrepreneurs enter the credit regime as implied by a reduction in $\bar{b}$. Gains on the extensive margin arise since external credit enables entrepreneurs to expand their businesses and produce more output. Moreover, gains on the intensive margin (within the credit regime) accrue since relaxing financial constraints limits the losses from financial contracts (lower ψ) and inefficient monitoring (lower χ) and improves the provision of credit (higher λ). When general equilibrium effects are considered, gains on both margins are likely to be smaller due to the increase in the equilibrium wage and interest rate. In fact, entrepreneurs living in the savings regime shrink the size of their businesses and produce less output after financial inclusion due to the general equilibrium effect. This implies that not everyone is better off with greater financial inclusion. In particular, entrepreneurs remaining in the savings regime would incur losses due to higher equilibrium factor prices. Some entrepreneurs in the credit regime can also lose if gains on the intensive margin are lower than the losses from the general equilibrium effects. We discuss the heterogeneous welfare effects in Subsection 6.4.

An important variation which is not captured in the GDP decomposition is the endogeneity of the steady-state distribution of wealth and talent. Equation (4.6) holds only for analyzing the immediate impact of financial inclusion as the distribution of wealth and talent $h(b, z)$ is assumed to be unchanged. However, in the new steady state after relaxing financial constraints, the endogenously determined joint distribution would also have a new shape. Therefore, the long-run effects of financial inclusion cannot be easily evaluated based on equation (4.6). We provide a numerical analysis in Subsection 6.3.

4.2.2 TFP Decomposition

In this subsection, we provide a TFP decomposition in the spirit of Jeong and Townsend (2007) by exploiting the equivalence between growth accounting by factors and growth accounting by regimes. Again, we consider relaxing the financial constraints from (ψ, λ, χ) to (ψ', λ', χ') with $\psi' \leq \psi$, $\lambda' \geq \lambda$, and $\chi' \leq \chi$. Our decomposition generalizes that of Jeong and Townsend (2007) since in our model the credit access ratio is endogenous and depends on the three micro-founded financial frictions. All the variables presented in this subsection are functions of (ψ, λ, χ), and the explicit dependence is omitted unless necessary.

We follow Buera and Shin (2013) and define the model-implied TFP as

$$TFP = \frac{Y}{K^{\alpha} L^{1-\alpha}}, \tag{4.7}$$

where Y is aggregate output, K is aggregate capital, and L is aggregate labor.

Growth Accounting by Factors By taking an approximation of the first difference of equation (4.7), we obtain

$$g_{TFP} = g_Y - \alpha g_K - (1-\alpha) g_L, \tag{4.8}$$

where g_x is the growth rate of variable x, i.e., $g_x = \dfrac{X(\psi', \lambda', \chi')}{X(\psi, \lambda, \chi)} - 1$.

Since the economy consists of two regimes, aggregate capital and labor are equal to the sum of capital and labor employed by entrepreneurs living in the two regimes separately. Denote $\overline{K}^s/\overline{L}^s$, and $\overline{K}^c/\overline{L}^c$ as the average capital/labor employed by entrepreneurs in the savings regime, and the credit regime, respectively. Denote p_c as the percent of entrepreneurs living in the credit regime and E as the total number of entrepreneurs. Therefore, aggregate capital and labor can be written as,

$$\begin{aligned} K &= E(1-p_c)\overline{K}^s + E p_c \overline{K}^c, \\ L &= E(1-p_c)\overline{L}^s + E p_c \overline{L}^c. \end{aligned} \tag{4.9}$$

From (4.9), the growth rates of aggregate factors g_K and g_L can be further decomposed into,

$$\begin{aligned} g_K &= g_E + (s_K^c - s_K^s)g_{p_c} + s_K^s g_{\overline{K}^s} + s_K^c g_{\overline{K}^c}, \\ g_L &= g_E + (s_L^c - s_L^s)g_{p_c} + s_L^s g_{\overline{L}^s} + s_L^c g_{\overline{L}^c}, \end{aligned} \tag{4.10}$$

where $s_K^c = \dfrac{E(1-p_c)\overline{K}^s}{2K}(\psi, \lambda, \chi) + \dfrac{E(1-p_c)\overline{K}^s}{2K}(\psi', \lambda', \chi')$ is the average fraction of capital employed by entrepreneurs in the credit regime before and after relaxing the financial constraints. The variables s_K^s, s_L^c, and s_L^s are defined in a similar way. Note that $s_K^s + s_K^c = 1$, and $s_L^s + s_L^c = 1$.

Substituting (4.10) into (4.8), we obtain

$$g_Y = g_{TFP} + g_E + \alpha(s_K^c - s_K^s)g_{p_c} + \alpha s_K^s g_{\overline{K}^s} + \alpha s_K^c g_{\overline{K}^c} + (1-\alpha)(s_L^c - s_L^s)g_{p_c} + (1-\alpha)s_L^s g_{\overline{L}^s} + (1-\alpha)s_L^c g_{\overline{L}^c}. \tag{4.11}$$

Growth Accounting by Regimes The economy's output Y is equal to the sum of output in each regime,

$$Y = E(1-p_c)\overline{Y}^s + Ep_c\overline{Y}^c, \tag{4.12}$$

where $\overline{Y}^s$ and $\overline{Y}^c$ are the average output produced by entrepreneurs in the savings and credit regimes, respectively.

Thus, the growth rate of output can be expressed as

$$g_Y = g_E + (s_Y^c - s_Y^s)g_{p_c} + s_Y^s g_{\overline{Y}^s} + s_Y^c g_{\overline{Y}^c}, \tag{4.13}$$

where $s_{Y^s} = \dfrac{Y^s}{2Y}(\psi, \lambda, \chi) + \dfrac{Y^s}{2Y}(\psi', \lambda', \chi')$ and $s_{Y^c} = \dfrac{Y^c}{2Y}(\psi, \lambda, \chi) + \dfrac{Y^c}{2Y}(\psi', \lambda', \chi')$ are the average fraction of output produced in the savings regime and credit regime, respectively.

Decomposition Formula Equating the two growth accounting identities (4.11) and (4.13), we obtain,

$$\begin{aligned} g_{TFP_{(\psi,\lambda,\chi)}\text{ to }_{(\psi',\lambda',\chi')}} &= \underbrace{(s_Y^c - s_Y^s - \alpha(s_K^c - s_K^s) - (1-\alpha)(s_L^c - s_L^s))g_{p_c}}_{\text{Between-regime shifting}} \\ &\quad + \underbrace{s_Y^s g_{\overline{Y}^s} - \alpha s_K^s g_{\overline{K}^s} - (1-\alpha)s_L^s g_{\overline{L}^s}}_{\text{Growth within savings regime}} \\ &\quad + \underbrace{s_Y^c g_{\overline{Y}^c} - \alpha s_K^c g_{\overline{K}^c} - (1-\alpha)s_L^c g_{\overline{L}^c}}_{\text{Growth within credit regime}} \end{aligned} \tag{4.14}$$

Therefore, the growth rate of TFP is decomposed into three terms. The first term captures growth generated by entrepreneurs who shift from the savings regime to the credit regime. In fact, the between-regime shifting effect can be considered as TFP gains on the extensive margin. The second and third term captures TFP growth within the savings regime and the credit regime, respectively. TFP growth within the credit regime can be considered as TFP gains on the intensive margin, and growth within the savings regime is caused by general equilibrium effects.

To evaluate and identify the macro, general equilibrium, and long-term impact of financial inclusion, a calibration and quantitative analysis is examined in the following sections.

5 Data and Calibration

We calibrate the model for six countries at various stages of economic development: three low-income countries (Uganda in 2005, Kenya in 2006, and Mozambique in 2006), and three emerging market economies (Malaysia in 2007, the Philippines in 2008 and Egypt in 2007). We use the data from World Bank Enterprise Surveys and World Development Indicators (WDI).[19]

In general, financial inclusion in low-income countries is more constrained compared with emerging market economies across different dimensions, as indicated by high collateral requirements, low share of firms with credit, and high borrowing costs (see Table 1). However, there is significant heterogeneity within country groups across these different dimensions. For example, access to the financial system, as measured by the share of firms with credit, is lower in Mozambique than in Uganda and Kenya, despite relatively lower collateral requirements and interest rate spreads. In the Philippines, collateral requirements are very high, while interest rate spreads are comparable to other emerging market economies in the sample.

Table 1: Overview of the Data

	Low-income countries			Emerging market economies		
	Uganda	Kenya	Mozambique	Malaysia	The Philippines	Egypt
Savings (% of GDP)	8	15.4	7.1	39	25.7	24.5
Collateral (% of loan)	173	120.8	92	64.6	238.4	85.5
Firms with credit (%)	17.2	25.4	14.2	60.4	33.2	17.4
Non-perfor. loan (%)	2.3	10.6	3.1	8.5	4.5	19.3
Interest rate spread	10.9	8.5	8.2	3.3	4.3	6.1
Overhead costs/assets	6.9	6.6	7.4	1.5	3.2	1.5
Top 5% emp. share	53.8	54.1	41.3	29.5	52.7	58.4
Top 10% emp. share	64.2	66.	55.8	46.3	65.7	72.7
Top 20% emp. share	74.6	81	71.9	63.5	79	85.9
Top 40% emp. share	86.4	93.2	87.2	84.1	90.8	95

We use standard values from the literature for some of the parameters. The one-year depreciation rate δ is set to be 0.06. Following Buera and Shin (2013), we choose the share of output going to the variable factors in the production function $1 - v$ to be 0.79, of which the share of capital α is

[19]The selection of countries is mainly driven by data availability. First and foremost, we need sufficient cross-section units to run our framework. The numbers of firms in our sample are 563 for Uganda, 781 for Kenya, 599 for Mozambique, 1115 for Malaysia, 1326 for the Philippines, and 996 for Egypt. Second, we consider relatively recent cases but exclude countries with financial turbulence around the year of the survey.

0.33. The probability that the offspring inherits the talent of her parents γ is assumed to be 0.894. The other parameters are estimated by matching the simulated moments to real data.

Table 2: Data, model and calibrated parameters

	Uganda			Kenya			Mozambique		
Target Moments	Data	Model	Parameter	Data	Model	Parameter	Data	Model	Parameter
Savings (% of GDP)	8	8	$\omega = 0.08$	15.4	15.4	$\omega = 0.15$	7.1	7.1	$\omega = 0.07$
Collateral (% of loan)	173	173	$\lambda = 1.58$	120.8	120.8	$\lambda = 1.83$	92	92	$\lambda = 2.09$
Firms with credit (%)	17.2	17.1	$\psi = 0.03$	25.4	25.1	$\psi = 0.07$	14.2	14.2	$\psi = 0.03$
Non-perfor. loan (%)	2.3	2.4	$p = 0.15$	10.6	5.8	$p = 0.17$	3.1	3.1	$p = 0.14$
Interest rate spread	10.9	8.9	$\chi = 0.90$	8.5	11.3	$\chi = 0.61$	8.2	11.2	$\chi = 0.95$
Overhead costs/assets	6.9	6.6	$\eta = 0.37$	6.6	6.5	$\eta = 0.45$	7.4	7.3	$\eta = 0.54$
Top 5% emp. share	53.8	52.9	$\theta = 4.80$	54.1	58.1	$\theta = 4.40$	41.3	46.9	$\theta = 6.00$
Top 10% emp. share	64.2	64.4		66.9	70.1		55.8	58.9	
Top 20% emp. share	74.6	74.7		81	80.5		71.9	69.1	
Top 40% emp. share	86.4	84.7		93.2	88.7		87.2	80.5	
	Malaysia			The Philippines			Egypt		
Target Moments	Data	Model	Parameter	Data	Model	Parameter	Data	Model	Parameter
Savings (% of GDP)	39	39	$\omega = 0.39$	25.7	25.7	$\omega = 0.26$	24.5	24.5	$\omega = 0.25$
Collateral (% of loan)	64.6	64.6	$\lambda = 2.56$	238.4	238.4	$\lambda = 1.42$	85.5	85.5	$\lambda = 2.17$
Firms with credit (%)	60.4	60.5	$\psi = 0.13$	33.2	33.0	$\psi = 0.07$	17.4	17.5	$\psi = 0.23$
Non-perfor. loan (%)	8.5	7.6	$p = 0.12$	4.5	3.8	$p = 0.11$	19.3	15.7	$p = 0.28$
Interest rate spread	3.3	5.8	$\chi = 0.16$	4.3	6.2	$\chi = 0.6$	6.1	8.0	$\chi = 0.08$
Overhead costs/assets	1.5	1.5	$\eta = 0.37$	3.2	3.1	$\eta = 0.27$	1.5	1.4	$\eta = 0.44$
Top 5% emp. share	29.5	34.5	$\theta = 6.80$	52.7	54.5	$\theta = 4.30$	58.4	62.0	$\theta = 4.25$
Top 10% emp. share	46.3	46.9		65.7	66.0		72.7	74.2	
Top 20% emp. share	63.5	61.5		79	77.0		85.9	83.5	
Top 40% emp. share	84.1	78.5		90.8	87.0		95	90.3	

Each generation is interpreted as one year as in Gine and Townsend (2004) and Jeong and Townsend (2008).[20] We match the gross savings rate, which measures the overall funds available for financial intermediation in a closed economy, in the data and the model to calibrate the optimal bequest rate, ω. We use the average value of collateral as a percentage of the loan to calibrate parameter λ, which captures the degree of financial frictions caused by limited commitment.[21]

The credit participation cost ψ, the intermediation cost χ, the probability of production failure p, the project recovery rate η, and the parameter governing the talent distribution θ are jointly calibrated to match the moments of the percent of firms with a line of credit, non-performing loans (NPLs) as a percentage of total loans, the interest rate spread, the bank overhead costs to assets ratio[22], and the employment share distribution (using four brackets of employment shares—top 5% / 10% /20% / 40%). Even though parameters ψ, χ, p, η, and θ affect the value of all these moments, and are jointly calibrated, each moment is primarily affected by particular parameters. Specifically,

[20] In Appendix B, we consider a model with forward-looking agents as a robustness check.

[21] Note that the definition of the loan-to-collateral ratio in the model is slightly different from that in the data. In our model, we assume that entrepreneurs deposit their wealth b in banks as collateral and borrow k^*. Hence, the loan-to-collateral ratio is $\lambda_{model} = \frac{k^*}{b}$. In the data, the loan-to-collateral ratio is measured as $\lambda_{data} = \frac{k^* - b}{b}$ (i.e., entrepreneurs put their own wealth into the project, and the wealth is used as collateral to borrow the difference $k^* - b$). The relationship between the two is $\lambda_{data} = \lambda_{model} - 1$.

[22] The bank overhead costs to assets ratio is obtained from World Bank Global Financial Development Database (GFDD). We compute its average value over the period 2000 – 2011. Greenwood, Sanchez and Wang (2013, Figure 6) show that this variable reflects monitoring efficiency in a cross-country analysis.

the moment of percent of firms with credit is mostly determined by the credit participation cost ψ. Increasing the value of ψ decreases the percent of firms with credit. The NPL ratio is mainly determined by parameter p. However, the relationship is non-monotonic for some parameter values. For example, when p increases, if entrepreneurs' leverage ratios were unchanged, NPLs should increase. However, a higher p may reduce leverage ratios as higher riskiness induce entrepreneurs to take less loans, which results in a lower NPL ratio. The employment share distribution is matched primarily by adjusting the value of parameter θ, which governs the shape of the entrepreneurial talent distribution. Parameter η is jointly identified with parameter χ to simultaneously match the bank overhead costs to assets ratio and the interest rate spread.

From Table 2, it is clear that the model performs well in terms of matching the macroeconomic moments. The percent of firms with credit and the bank overhead costs to assets ratio generated by the model are almost exactly matched with those in the data for all six countries. Both NPLs and interest rate spreads are matched well, although some countries have high NPLs but relatively low interest rate spreads (e.g. Maylasia and Egypt) while other countries have low NPLs and high interest rate spreads (e.g. Uganda and Mozambique). The employment share distribution is also captured, but in general the model tends to generate more large firms compared to the data.

The linkages between different characteristics of an economy and financial inclusion are complex. For example, it might seem surprising that the calibrated credit participation cost ψ, in general, is lower in low-income countries despite their lower credit access ratios. This is because ψ is not the only factor determining credit access. In fact, both λ and χ affect the credit access ratio in the model—higher λ and lower χ increase credit access in emerging market economies, whose partial equilibrium effects are summarized in Theorem 1. Moreover, these countries have higher savings rates (higher ω), which implies that agents transfer more wealth to the next generation. In this case, the credit participation cost is a relatively smaller proportion of agents' wealth in emerging market economies. Therefore, it is less binding, as reflected in the high credit access ratio. In the next section, we analyze the macroeconomic implications of financial inclusion and identify the role that country characteristics play in the process.

6 Quantitative Analysis

In our quantitative analysis, we first evaluate the impact of relaxing different financial constraints individually for the countries in our sample. Then we take the Phillipines as an example to evaluate the interaction effect among the three financial constraints. Finally, we decompose GDP and TFP and provide a welfare analysis.

6.1 Evaluation of Policy Options

This subsection analyzes the policy implications of promoting financial inclusion across these three dimensions for the countries in our sample. Specifically, we focus on changes across the steady states of the economy when these parameters change.[23] Figures 5 – 10 below present the simulation results when each of the three financial parameters changes separately (on the horizontal axis).

6.1.1 Reducing the Participation Cost

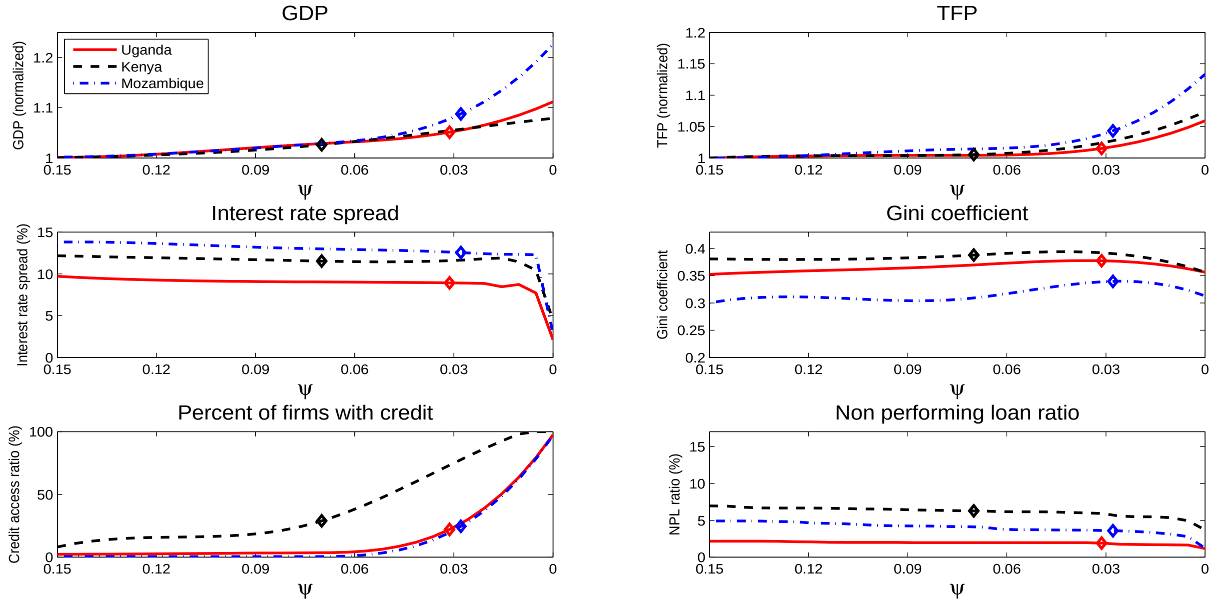
Figures 5 – 6 present the impact of a decline in the credit participation cost ψ from 0.15 to 0 (moving from left to right). A decrease in the participation cost pushes up GDP through its positive impact on investment for two reasons. First, a lower credit participation cost enables more entrepreneurs to have access to credit. Second, less funds are wasted in unproductive contract negotiation. Both tend to increase entrepreneurs' investment in production. TFP increases as capital is more efficiently allocated among entrepreneurs.

The average interest rate spread is stable when ψ is high, but eventually decreases in some countries (Uganda, Mozambique, Mozambique and the Philippines). A smaller ψ enables some of the constrained workers to become entrepreneurs. These entrepreneurs are severely wealth constrained, and therefore choose high leverage ratios, driving the average interest rate spread up.[24]

As financial inclusion increases, income inequality (the Gini coefficient in our simulation) first increases and then decreases in low-income countries, consistent with the Kuznets' hypothesis. This is because when ψ decreases from a particularly high value, it only enables a very small number of constrained workers to become entrepreneurs. As shown in Figure 5, the percent of firms with credit is almost unchanged for high values of ψ. However, the effect on the incumbent entrepreneurs is large since it reduces their contracting costs, thus allowing them to invest more capital in production. These entrepreneurs make higher profits, pushing up income inequality. If ψ decreases further (all the way to zero), it becomes disproportionately more beneficial for constrained workers and entrepreneurs without access to credit. This enables relatively poorer agents to earn higher income, driving down the Gini coefficient.

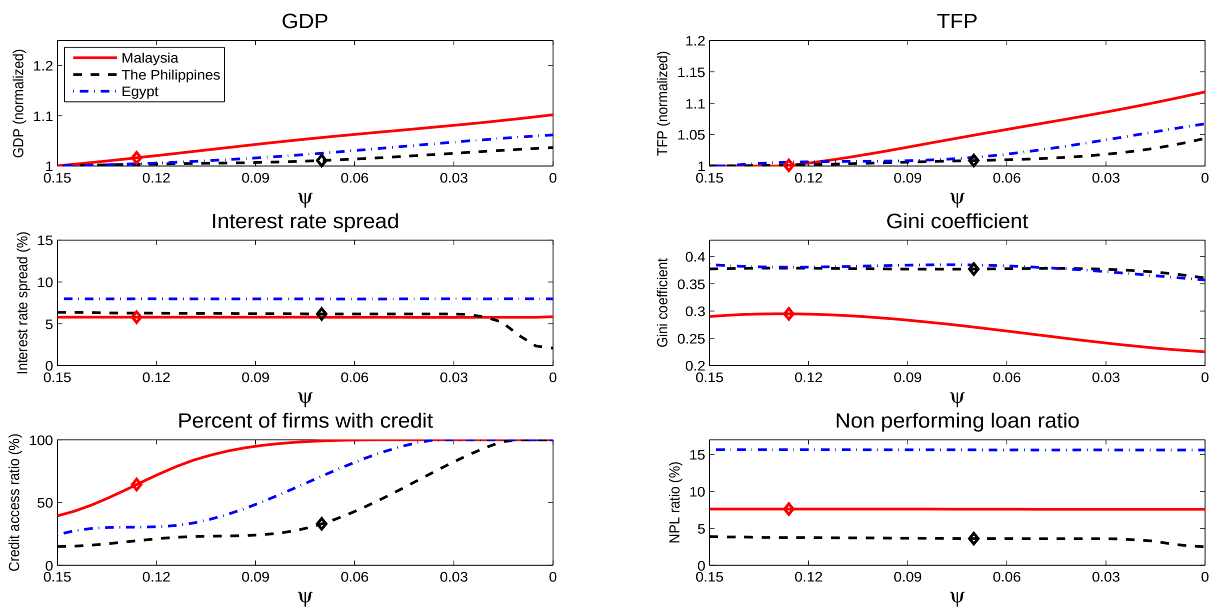
[23] Note that it takes time for the economy to transition from one steady state to another when these parameters change. The models' transitional dynamics are also computable. However, we only report the simulation results in steady states because focusing on the transitional dynamics could be misleading for at least two reasons: (1) The transition is rapid at the beginning but becomes slower when the economy is approaching the steady state. This is inconsistent with reality, where the impact of financial reforms happen gradually, or at least the immediate impact is not significant; and (2) Numerical errors are large relative to those in the steady state, possibly leading to overshooting in some variables if parameters are adjusted a lot. These two problems associated with computing transitional dynamics exist for all quantitative macroeconomic models, although the first problem could be mitigated to some extent if agents were modeled as forward-looking (e.g. Buera and Shin, 2013, and our robustness check).

[24] A decrease in ψ also has a countervailing effect on interest rate spreads in our model. As entrepreneurs need to pay less to get credit, they become richer and tend to deleverage, resulting in a lower interest rate spread. However, this effect is secondary.



Note: The solid line represents Uganda, the dashed line represents Kenya, and the dash-dotted line represents Mozambique. The circle on each line represents the country's position in the survey date (i.e. at the calibrated parameter values).

Figure 5: Comparative statics: Credit participation cost—low-income countries.



Note: The solid line represents Malaysia, the dashed line represents the Philippines, and the dash-dotted line represents Egypt. The circle on each line represents the country's position in the survey date (i.e. at the calibrated parameter values).

Figure 6: Comparative statics: Credit participation cost—emerging market economies.

By contrast, in emerging market economies, this Kuznets' pattern is not observed. The reason is that at $\psi = 0.15$, financial systems in these economies are already highly developed compared to low-income countries. In other words, emerging market economies are already in the "second stage" of development.[25] A decrease in ψ unambiguously leads to a lower Gini coefficient in emerging market economies, such as Malaysia. Since ψ is a fixed cost, a decrease in ψ benefits poor entrepreneurs disproportionately as this constitutes a larger proportion of their wealth. In the Philippines and Egypt, the decline in inequality is less noticeable, reflecting other binding constraints to financial inclusion.

6.1.2 Relaxing the Borrowing Constraint

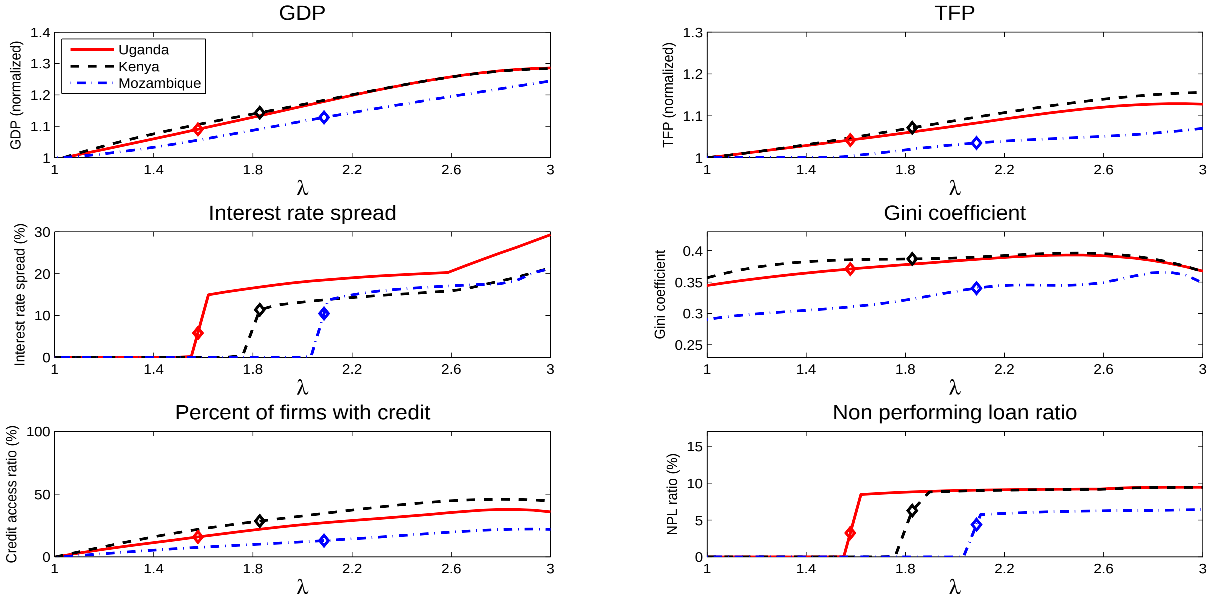
In Figures 7 – 8, we vary the borrowing constraint parameter λ from 1 to 3. Following the relaxation of the borrowing constraint, GDP increases in all countries. However, the responsiveness of output is highly dependent on the economy's savings rate. In low-income countries, GDP is typically more responsive as agents' production relies heavily on external financing due to small transfers across periods (low savings rates). This suggests that the borrowing constraint is one of the major obstacles to economic development for low-income countries in our sample. In the Philippines, GDP also responds well to the relaxation of the borrowing constraint; however, the reason for this is different from that for low-income countries. Financial access is moderate in the Philippines, but interest spreads are low and savings rates are high. Therefore, the relaxation of the borrowing constraint unlocks financial resources, leading to a significant increase in GDP.

As λ increases, TFP increases, implying a more efficient resource allocation across entrepreneurs. A relaxation of the borrowing constraint benefits talented entrepreneurs disproportionately as they often desire to operate businesses at a larger scale than untalented entrepreneurs (i.e., both $\tilde{k}^l(z)$ and $\tilde{k}^h(z)$ increase in z). Relaxing the borrowing constraint allows all entrepreneurs to borrow more, but on average untalented ones do not borrow as much because their smaller (maximum) business scale may have already been achieved. As a result, more talented entrepreneurs enter the credit regime and expand their scale of operations, driving up TFP.

The interest rate spread increases in this scenario. The spread is zero when λ is low, because entrepreneurs' leverage is low—no default happens even when production fails. As λ increases above a threshold, agents leverage more, the share of NPLs increases, and the interest rate spread starts increasing. Note that, in general, low-income countries have higher interest rate spreads relative to emerging market economies due to higher intermediation costs.

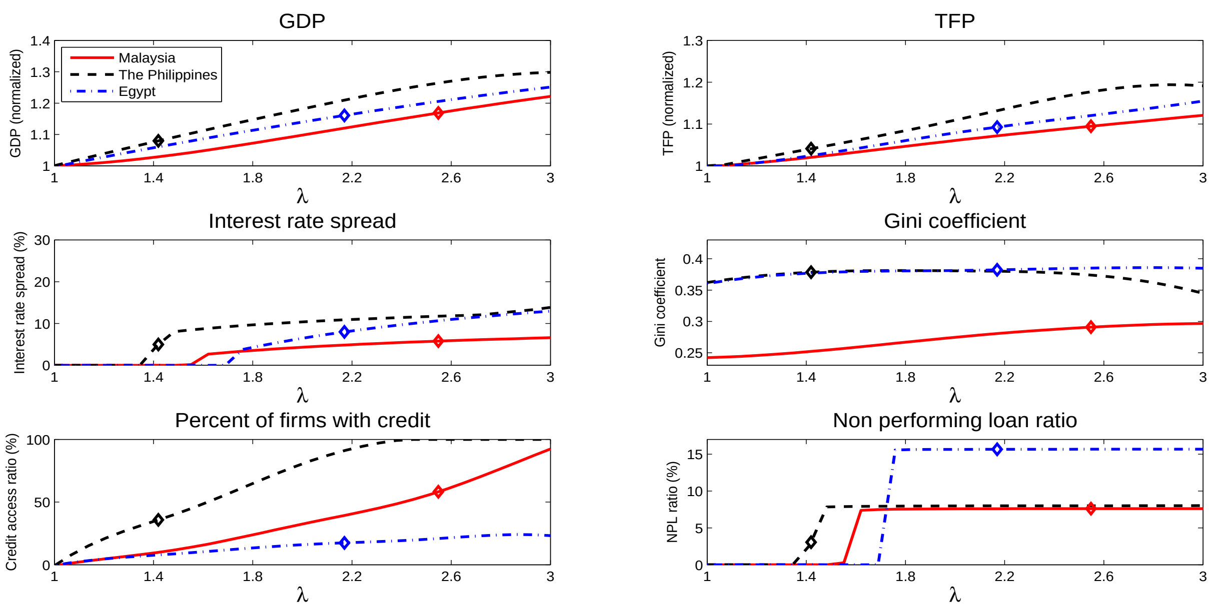
In terms of inequality, the Kuznets pattern is again observed for low-income countries. As λ increases, talented entrepreneurs can leverage more and increase their profits, which drives up the

[25] As reflected in Figures 5 – 6 , at $\psi = 0.15$, the percent of firms with credit is about 50% in Malaysia while it is close to zero in Uganda. Identified by the circle on the blue solid line in Figure 5, Uganda in 2005 was about to move from the initial stage of development (in the Kuznets' sense).



Note: The solid line represents Uganda, the dashed line represents Kenya, and the dash-dotted line represents Mozambique. The circle on each line represents the country's position in the survey date (i.e. at the calibrated parameter values).

Figure 7: Comparative statics: Borrowing constraint—low-income countries.



Note: The solid line represents Malaysia, the dashed line represents the Philippines, and the dash-dotted line represents Egypt. The circle on each line represents the country's position in the survey date (i.e. at the calibrated parameter values).

Figure 8: Comparative statics: Borrowing constraint—emerging market economies.

Gini coefficient. In low-income countries, the savings rate is low. As a result, external credit is limited and the interest rate increases by more, the easier the borrowing constraint is. As λ becomes larger, the sharp increase in the interest rate shrinks entrepreneurs' profits, leading to a lower Gini coefficient.

Relaxing the borrowing constraint provides more external credit to entrepreneurs once they pay the participation cost. This induces more entrepreneurs to join the credit regime. However, NPLs also increase. This occurs as a relaxation of the borrowing constraint opens up the doors for small new entrants who tend to be more leveraged.

6.1.3 Increasing Intermediation Efficiency

In Figure 9 – Figure 10, we vary the monitoring cost χ from 1.2 to 0 to reflect financial inclusion from an intermediation efficiency angle. When χ decreases, the response of GDP varies across countries. In some countries (Uganda, Mozambique and the Philippines), GDP is not responsive as a lower intermediation cost only benefits highly-leveraged entrepreneurs, which are few due to the low credit access ratio.

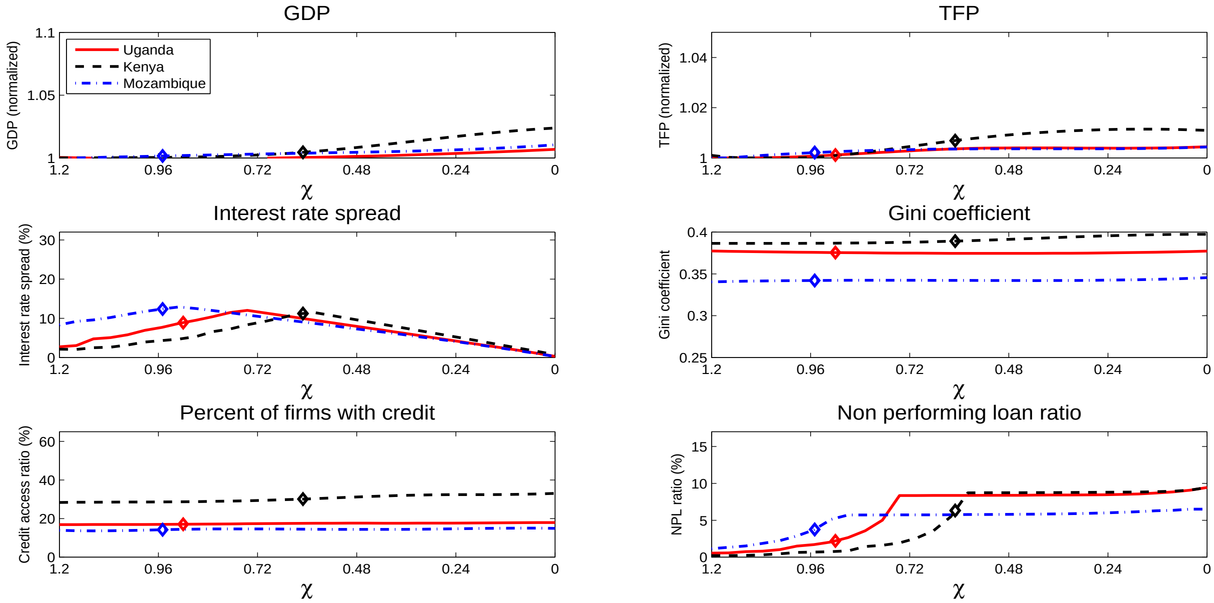
TFP increases (but only slightly) as a lower intermediation cost facilitates the allocation of capital to efficient entrepreneurs. The interest rate spread monotonically declines in Malaysia, but displays an inverted V-shape in other countries. Two opposing forces are in effect here. First, the decline in the net lending rate induces entrepreneurs to increase leverage because it reduces the cost of capital for risky firms, pushing up the share of NPLs. This tends to increase the endogenous interest rate spread. Second, a lower intermediation cost decreases the interest spread through its pass-through effect. Whether the interest rate spread increases or decreases depends on which effect dominates.

The Gini coefficient increases as efficient intermediation disproportionately benefits highly-leveraged entrepreneurs (who already have higher income than others).[26] Moreover, a lower intermediation cost induces more agents to borrow, hence increasing the percent of firms with credit.

6.1.4 Impact on GDP and Inequality: A Numerical Comparison

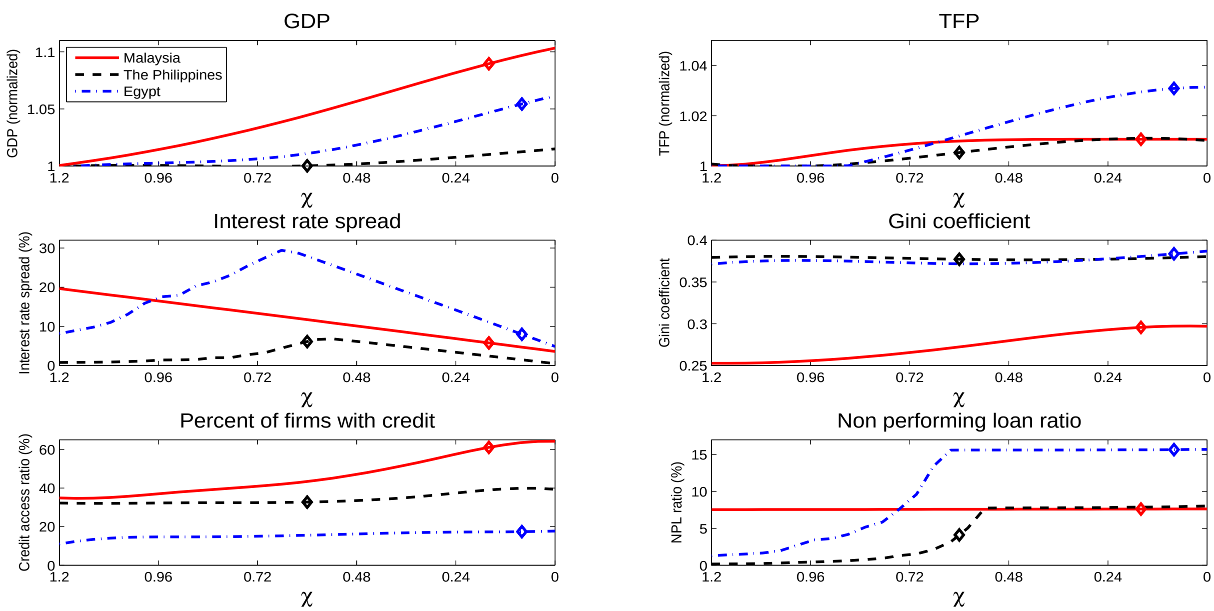
Figures 5 – 10 suggest that the economic implications of financial inclusion policies depend on the source of frictions. In this subsection, we zoom-in on a numerical comparison of the marginal responses of GDP and inequality. The numbers in Table 3 are calculated as the difference between the current state of the country (shown with circles in Figures 5 – 10) and the eventual steady state when the economy's credit to investment ratio is increased by one percentage point.

[26] There is only a slight increase (almost invisible on the figure) in the Gini coefficient of Uganda, Mozambique, and the Philippines, as the model suggests that the intermediation cost is not a binding constraint in these countries.



Note: The solid line represents Uganda, the dashed line represents Kenya, and the dash-dotted line represents Mozambique. The circle on each line represents the country's position in the survey date (i.e. at the calibrated parameter values).

Figure 9: Comparative statics: Intermediation cost—low-income countries.



Note: The solid line represents Malaysia, the dashed line represents the Philippines, and the dash-dotted line represents Egypt. The circle on each line represents the country's position in the survey date (i.e. at the calibrated parameter values).

Figure 10: Comparative statics: Intermediation cost—emerging market economies.

As before, although financial inclusion brings an increase in GDP and TFP in all cases, its impact on inequality varies. The impact on the Gini coefficient can be positive or negative for a reduction in the credit participation cost, depending on country-specific characteristics.

Moreover, in line with the discussion above, the numbers highlight that the form of financial inclusion and country characteristics matter in how the economies respond. For example, Uganda's GDP responds more if the increase in credit to investment ratio comes from a lower participation cost. However, Egypt's GDP responds more to relaxing the borrowing constraint; while the other countries are more responsive to a lower intermediation cost.[27]

Table 3: The impact of financial inclusion of various forms on GDP per capita, TFP and income inequality.

	Participation cost ψ			Borrowing constraint λ			Intermediation cost χ		
	GDP(%)	TFP(%)	Gini	GDP(%)	TFP(%)	Gini	GDP(%)	TFP(%)	Gini
Uganda	0.40	0.28	-0.0007	0.35	0.20	0.0007	0.03	0.14	-0.0006
Kenya	0.67	0.40	0.0033	0.28	0.22	0.0001	0.07	0.10	0.0004
Mozambique	0.38	0.28	0.0002	0.29	0.13	0.0011	0.38	0.10	-0.0001
Malaysia	0.38	0.37	-0.0005	0.52	0.32	0.0010	1.10	0.00	0.0015
The Philippines	0.28	0.16	0.0006	0.20	0.15	0.0002	0.04	0.08	-0.0003
Egypt	0.26	1.26	-0.0093	0.46	0.35	0.0003	0.69*	0.02*	0.0033*

Note: We consider the change of parameters that increases the credit to investment ratio by 1%. In cases marked with *, we report the change in GDP, TFP, and Gini when parameter χ is reduced to zero. This is because in these cases, even if parameter χ is reduced to zero, the increase in the credit to investment ratio is still less than 1%.

How far are these countries from the world best financial sector technology in terms of these three financial parameters? Which country is most underdeveloped along which dimension? To shed light on these questions, we show a numerical comparison for changes in GDP, TFP and the Gini coefficient when the six countries adopt the best-possible intermediation technology. Obviously, the best possible value for the credit participation cost and monitoring cost are zero ($\psi = \chi = 0$). Among the 127 countries in enterprise surveys, we consider countries that require the lowest amount of collateral (Germany, Spain, and Portugal). The average amount of collateral required as a percent of loans in these countries is about 50% ($\lambda = 3$), which is regarded as the best possible borrowing constraint.

Table 4 shows the simulation results when one of the financial parameters is equal to the world frontier value. The increase in GDP is largest when the borrowing constraint is relaxed in Uganda, Kenya, the Philippines and Egypt, implying that the financial sector in these countries is facing disproportionately higher collateral requirements. By contrast, Mozambique and Malaysia's GDP are more responsive to a decrease in the credit participation cost, indicating that limited credit availability or low financial access is the major obstacle. Moreover, reducing the credit participation cost leads to a uniform increase in TFP and decrease in the Gini coefficient in all countries for reasons

[27] Using the credit to investment ratio might bias the results on the effectiveness of different sources of financial inclusion since the credit to investment ratio itself is more responsive to some factors (e.g. λ), and significantly less responsive to other factors (e.g. χ). Therefore, the impact of λ is likely to be underestimated, while the impact of χ is likely to be overestimated.

discussed above, while relaxing the borrowing constraint increases TFP, but has an ambiguous impact on income inequality. Not surprisingly, adopting the most efficient intermediation technology ($\chi = 0$) does not boost GDP significantly. However, this does not imply that the intermediation cost is not crucial in terms of financial inclusion. As we show in Theorem (1), (2) and Subsection 6.2, there exist rich interactions among these parameters: inefficient intermediation will dampen the responsiveness of GDP to a lower credit participation cost and a relaxed borrowing constraint, or even block these channels.

Table 4: The impact of financial inclusion of various forms on GDP per capita, TFP and income inequality.

	Participation cost ψ			Borrowing constraint λ			Intermediation cost χ		
	GDP(%)	TFP(%)	Gini	GDP(%)	TFP(%)	Gini	GDP(%)	TFP(%)	Gini
Uganda	5.77	5.67	-0.0210	17.94	10.41	-0.0034	0.74	0.42	0.0018
Kenya	5.16	6.50	-0.0314	12.28	9.30	-0.0203	1.93	0.74	0.0082
Mozambique	12.72	10.16	-0.0267	10.30	4.83	0.0217	0.88	0.32	0.0033
Malaysia	8.44	10.94	-0.0696	4.52	2.85	0.0059	1.26	0.00	0.0013
The Philippines	2.56	3.40	-0.0165	20.21	16.45	-0.0336	1.48	0.58	0.0033
Egypt	7.04	11.31	-0.0590	7.78	6.61	0.0026	0.69	0.02	0.0033

Note: In all cases, we consider financial inclusion that moves the country to world financial sector frontier in one of the three parameters.

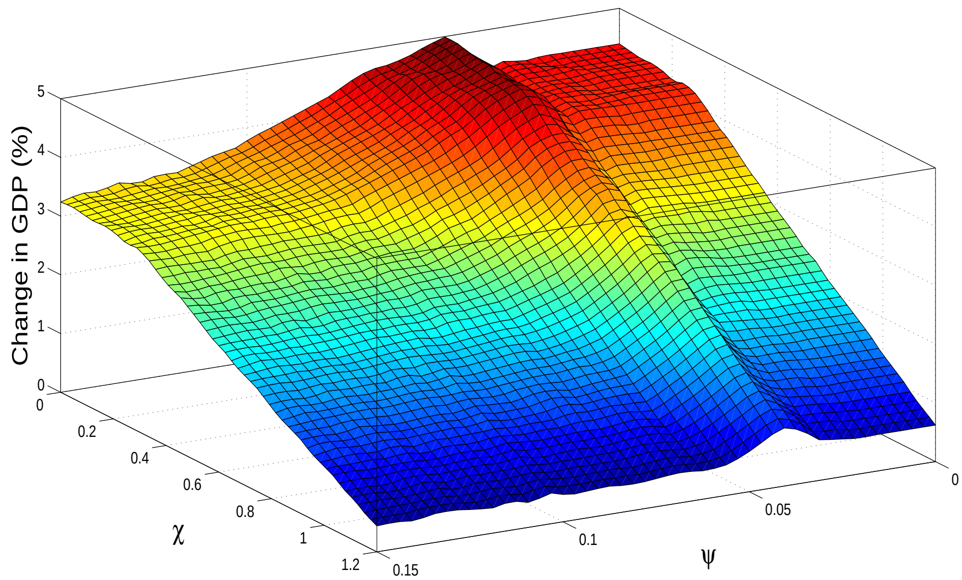
6.2 Interactions among the Three Financial Constraints

In Sections 6.1.1 – 6.1.3, we have shown that policies targeting different financial parameters have differential effects on macroeconomic aggregates. Moreover, the effects vary across countries depending on how country-specific economic characteristics interact with financial sector characteristics. In this subsection, we shed light on how the three financial sector parameters interact with each other and examine its implication on the macroeconomy and financial policies.

We take a specific country—the Philippines—and study the change in GDP per capita following a relaxation of the borrowing constraint (i.e. an increase in parameter λ).[28] In particular, we relax the borrowing constraint by 20%, and compare the increase in GDP relative to the previous state (i.e. before relaxing the borrowing constraint) for different levels of the credit participation cost ψ and the intermediation cost χ. Figure 11 shows that the relative change in GDP following an increase in λ depends on the two costs, ψ and χ. When χ increases, the increase in GDP becomes smaller for all ψ. This is because relaxing the borrowing constraint increases GDP by providing more credit to entrepreneurs. However, this channel is partially blocked if the intermediation cost is very high, due to the complementarity on the intensive margin as shown in Theorem 2. A higher intermediation cost restricts entrepreneurs from borrowing more as they want to keep low leverage ratios to avoid being monitored. This dampens the GDP-boosting effect that arises from a relaxation of the borrowing constraint. If the intermediation cost is too high, relaxing the

[28]Using other countries' calibrated parameters does not change the qualitative results we emphasize.

borrowing constraint would be futile as all entrepreneurs prefer to stay with low leverage ratios to avoid paying the monitoring cost.



Note: The horizontal axes refer to the intermediation cost χ and the credit participation cost ψ; The vertical axis refers to the relative change in GDP when the borrowing constraint λ is relaxed by 20%. The calibrated parameters for the Philippines are used for this study.

Figure 11: The increase in relative GDP per capita when the borrowing constraint is relaxed by 20% for different credit participation costs and intermediation costs.

However, the change in GDP is non-monotonic when ψ increases. The change in GDP stays almost constant for low values of ψ ($\psi < 0.03$); it is increasing in ψ when ψ lies between 0.03 and 0.04, and is decreasing for large values of ψ ($\psi > 0.04$). This non-monotonic pattern results from the two channels through which relaxing the borrowing constraint impacts GDP (as emphasized in Theorem 1 and 2). On the one hand, it enables agents in the credit regime to borrow more (the intensive margin). On the other hand, it induces more agents to join the credit regime, as a lower borrowing constraint increases the benefit of obtaining a credit contract (the extensive margin). Gains on both the intensive and extensive margins depend on the fraction of agents in the credit regime. A decrease in ψ promotes financial inclusion, increasing gains on the intensive margin (i.e. there is complementarity on the intensive margin between ψ and λ as shown in Theorem 2). However, it decreases gains on the extensive margin as relaxing the borrowing constraint has less of an impact on increasing the credit access ratio when this ratio is already high (i.e. there is a substitution effect on the extensive margin between ψ and λ as shown in Theorem 1). Therefore, as ψ decreases, change in GDP first increases and then decreases. The change in GDP, however, stays almost constant for low values of ψ. This is because the credit access ratio is about 100% when $\psi < 0.03$ (see Figure 6), so that further reducing ψ has no impact on gains accruing on both margins.

This exercise suggests that the effectiveness of financial inclusion policies depends crucially on the underlying financial sector characteristics within an economy. Relaxing the borrowing constraint is less effective if the intermediation cost is high, which is partially reflected in a high interest rate spread. The impact of relaxing the borrowing constraint also depends on the credit access ratio, although the relationship is not as clear-cut because of the coexistence of the two margins. The exercise also suggests that financial inclusion policies can be used in a complementary way in order to be more effective. For example, reducing the intermediation cost not only directly boosts GDP, but also amplifies the effect of relaxing the borrowing constraint. However, simultaneously reducing the participation cost and relaxing the borrowing constraint may be partially substitutable, as both policies increase GDP by promoting credit access. The optimal mix of policies thus depends on the underlying financial sector parameters and country-specific characteristics.

6.3 Decomposition of GDP and TFP

Still taking the Philippines as an example, we conduct a GDP decomposition using equation (4.6) and the endogenous steady-state distributions of wealth and talent. Table 5 reveals that relaxing different financial constraints promote GDP through different channels. In partial equilibrium, about two third of the increase in GDP from reducing the credit participation cost arises from gains on the extensive margin. However, most of the increase in GDP due to either relaxing the borrowing constraint or reducing the intermediation cost arises from gains on the intensive margin. This is intuitive since the credit participation cost can be considered as an ex-ante cost for credit while the borrowing constraint and the intermediation cost are both ex-post costs. The increase in GDP in general equilibrium is much smaller than that in partial equilibrium due to higher wages and interest rates. Notably, in general equilibrium, when the credit participation cost is lowered, there are losses on the intensive margin. This is because entrepreneurs who are already in the credit regime do not benefit much but the increase in wages and interest rates depresses their profits substantially. Finally, entrepreneurs in the savings regime also incur losses in general equilibrium due to the spill-over effects of equilibrium factor prices.

Next we conduct a TFP decomposition for the Philippines using equation (4.14). As shown in Table 6, most of the increase in TFP arises from the between-regime shifting effect regardless of which financial constraint is relaxed. Moreover, there are large TFP losses in the savings regime as talented entrepreneurs enter the credit regime when financial constraints are relaxed. In partial equilibrium, the qualitative results do not change significantly, but the magnitude of each TFP component is amplified while the increase in the economy's aggregate TFP is smaller. In fact, in general equilibrium, the increase in the equilibrium interest rate and wage drives out less talented entrepreneurs, hence increasing average productivity. Therefore, since prices are fixed in partial

Table 5: GDP decomposition.

		General Equilibrium		Partial Equilibrium	
		GDP(%)	Contribution(%)	GDP(%)	Contribution(%)
ψ	Extensive margin	8.94	348.96	3.69	68.08
	Intensive margin	-6.32	-246.73	1.73	31.92
	Savings regime	-0.06	-2.23	0	0
	Total	2.56	100	5.42	100
λ	Extensive margin	2.75	13.59	9.68	19.61
	Intensive margin	24.68	122.13	39.7	80.39
	Savings regime	-7.22	-35.72	0	0
	Total	20.21	100	49.38	100
χ	Extensive margin	0.09	5.76	0.09	4.27
	Intensive margin	1.79	120.70	2.04	95.73
	Savings regime	-0.39	-26.46	0	0
	Total	1.48	100	2.13	100

Note: In all cases, we consider financial inclusion that moves the Philippines to the world financial sector frontier in one of the three parameters.

equilibrium, the increase in TFP tends to be smaller.[29]

Table 6: TFP decomposition.

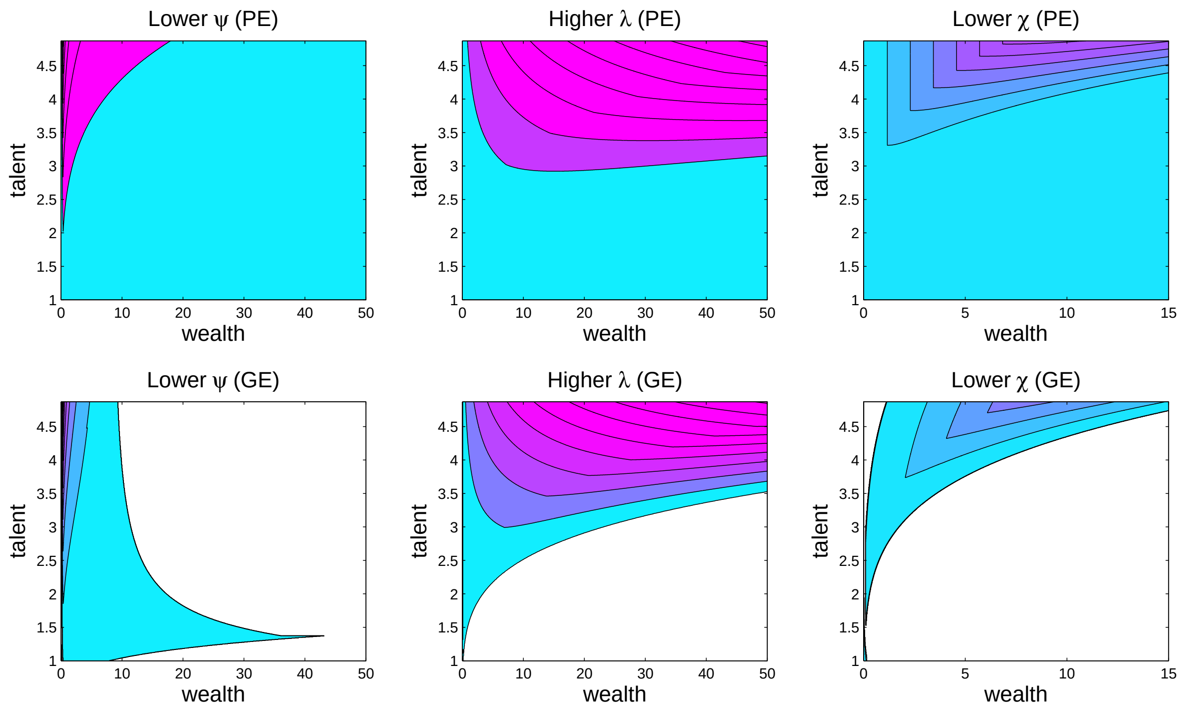
		General Equilibrium		Partial Equilibrium	
		TFP(%)	Contribution(%)	TFP(%)	Contribution(%)
ψ	Between-regime shifting	7.68	226.02	15.32	471.36
	Credit regime	-0.85	-25.05	-1.69	-51.95
	Savings regime	-3.43	-100.97	-10.38	-319.41
	Total	3.40	100	3.25	100
λ	Between-regime shifting	22.53	136.95	29.61	296.11
	Credit regime	4.16	25.27	-1.04	-104.05
	Savings regime	-10.24	-62.22	-12.56	-125.60
	Total	16.45	100	16.01	100
χ	Between-regime shifting	0.78	133.98	0.81	142.76
	Credit regime	-0.19	-32.36	-0.22	-38.80
	Savings regime	-0.01	-1.62	-0.02	-3.96
	Total	0.58	100	0.57	100

Note: In all cases, we consider financial inclusion that moves the Philippines to the world financial sector frontier in one of the three parameters.

6.4 Welfare Analysis

Financial inclusion engenders growth in GDP, however, not all agents are necessarily better off. In this subsection, we investigate the heterogeneous welfare redistribution effects following different financial inclusion policies. In particular, we quantify the amount of consumption change for different agents (endowed with different wealth and talent) when one of the financial sector parameters (ψ, λ, χ) changes. In Figure 12, we present partial equilibrium (top three panels) and general equilibrium (bottom three panels) results separately to highlight their differences. A comparison suggests that the changes in equilibrium interest rates and wages are sources of welfare losses for some agents. That is, if interest rates and wages were fixed, all agents would gain following financial inclusion.

[29]Quantitatively, the difference in TFP between partial equilibrium and general equilibrium is smaller compared with that in Buera, Kaboski and Shin (2012), which is due to myopic savings rate assumed here.



Note: The horizontal and vertical axes refer to wealth and talent, respectively. Consumption gains are reflected by the difference in shades of color—gains are low for light areas (white areas incur losses). Panels in the first row are partial equilibrium results when interest rates and wages are fixed; panels in the second row are general equilibrium results. The left, middle, and right columns represent the change of ψ, λ, and χ, respectively.

Figure 12: The impact of financial inclusion of various forms on welfare redistribution.

The left-most panels show the change in consumption when the credit participation cost ψ declines. Agents in the white areas experience a reduction in utility after financial inclusion. This is because a reduction in ψ enables more entrepreneurs to borrow, driving up equilibrium wages and interest rates. Wealthier agents lose as they benefit less from a lower credit participation cost and suffer more from the ensuing increase in wages and interest rates. Interestingly, the boundary line is not monotonic. As talent increases, the threshold level of wealth beyond which agents lose first increases then decreases. To understand this pattern we compare agents around the boundary line. We find that the lower part of the boundary line (talent < 1.5) separates agents who use external funds from those who do not. Agents to the right of the boundary line are sufficiently wealthy to self finance production. These agents do not demand external credit, so a reduction in the participation cost does not benefit them. However, because of the increase in wages and interest rates, these agents make lower profits when the participation cost declines. The threshold wealth level is increasing in talent when talent is below 1.5 because talented agents have a higher demand for capital, and therefore need to have higher wealth in order to self finance production. By contrast, when talent is above 1.5 and increases further, the threshold level of wealth decreases. This is because in our model, talented entrepreneurs disproportionately demand more labor than capital (see the optimal labor decision in Proposition 1), therefore they suffer more from the increase in wages. Since labor demand is also increasing in wealth, as talent increases, the marginal gainer should have lower wealth to mitigate the wage effect. Notice that the biggest winner after a reduction in the credit participation cost lies in the upper left corner. These agents are poor but very talented. The reduction in the participation cost enables them to have access to external credit, allowing them to expand businesses and increase their profits.

The middle two panels present the consumption change following a relaxation of the borrowing constraint. In this case, untalented entrepreneurs whose demand for credit is low lose. They incur consumption losses because they do not benefit as much from the relaxation of the borrowing constraint due to their low credit demand. Instead, they suffer from the increase in wages and interest rates. The biggest winners are talented and wealthy agents. As credit is proportional to wealth, relaxing the borrowing constraint enables wealthier agents to receive more funds, increasing their profits. Note that if talented and wealthy agents are not financially constrained, these agents will actually lose due to the general equilibrium effect. However, in our calibration, a severe credit constraint is observed for almost all agents due to low savings rates and the finite horizon framework.

The right-most panels show the consumption change following a decrease in the intermediation cost χ. The biggest winners are the most talented agents with moderate amounts of wealth. Intuitively, talented agents employ more capital, hence a reduction in the intermediation cost reduces their costs of production by more. Note that the biggest winners are not the wealthiest agents, because they already have sufficient internal funds, and have low demand for credit. Agents in the two white areas both experience a decrease in their income, but for different reasons. Agents in the upper-left area are talented but poor, and operate their businesses at the maximum leverage ratio. Hence, they benefit from the decrease in the intermediation cost. However, because their demand for capital is low, the benefit from the lower cost of capital is smaller than the increased cost of labor wages.[30] Agents in the lower-right area lose because they operate businesses with low leverage ratios (not being monitored), hence they do not receive benefits from a lower intermediation cost but suffer from the increase in labor costs.

7 Conclusion

We develop a tractable micro-founded general equilibrium model with heterogeneous agents to analyze the implications of financial inclusion policies on GDP and inequality in developing countries. In particular, we focus on three specific dimensions of greater financial inclusion: access (as measured by the size of credit participation costs), depth (as measured by the size of borrowing constraints resulting from limited commitment), and intermediation efficiency (as measured by the size of

[30] According to the labor demand function in Proposition 1, the capital/labor ratio is increasing in wealth. Therefore, poor agents benefit less from a reduction in the cost of intermediation.

interest rate spreads, reflecting default risk and asymmetric information).

We show theoretically that relaxing each constraint can have a differential quantitative impact on individuals' output and their credit access conditions. More importantly, there exist rich interactions among these financial constraints. In partial equilibrium, we prove that relaxing all constraints simultaneously can have complementary effects of increasing GDP on the intensive margin. However, these constraints are substitutes on the extensive margin. Such interaction effects in general equilibrium are confirmed in our quantitative analysis. Moreover, we find that the sources of GDP gains vary across financial policies. When the participation cost is reduced, GDP gains arise from the extensive margin. However, most of the GDP gains are from the intensive margin when the borrowing constraint and the intermediation cost are reduced. Our TFP decomposition indicates that irrespective of the sources of financial inclusion, most of the TFP gains are captured by a between-regime shifting effect, namely, talented entrepreneurs enter the credit regime and expand their businesses.

Using analytical and numerical methods, we calibrate the model for six countries—Uganda, Kenya, Mozambique, Malaysia, the Philippines, and Egypt. While our simulation results are intended to be illustrative, they indicate that relaxing various financial sector frictions may affect GDP and inequality in different ways. Our findings suggest that country-specific characteristics play a central role in determining the impacts, interactions, and trade-offs among policies. Thus, understanding the specific constraints generating the lack of financial inclusion in an economy is critical for tailoring policy recommendations.

A defining feature of our model is having three different types of financial frictions, limited participation, limited commitment, and asymmetric information (or costly state verification) all embedded in a unified framework. Most of the recent literature only considers one. We emphasize that the micro-foundations for each of the three financial frictions included in this paper are very different. Moreover, even within a given economy, there exists evidence that individuals face different types of financial frictions depending on location (Paulson, Townsend and Karaivanov, 2006; Ahlin and Townsend, 2007; Karaivanov and Townsend, 2014). Thus a quantitative framework that incorporates multiple types of financial frictions is needed to capture the multifaceted nature of financial systems. Even at the aggregate level, alleviating different sources of financial frictions can have differential impacts, either qualitatively or quantitatively, on key macro variables. Incorporating these frictions altogether in a unified framework enables us to identify the most binding constraints that hinder financial inclusion within an economy and uncover their interactions. Moreover, multiple frictions are necessary ingredients to match the data for a wide set of countries. If the only financial friction is limited commitment, relaxing it increases the credit access ratio, interest rate spread, and non-performing loans in tandem. However, in the data, the three moments are not perfectly positively correlated across countries. For example, Uganda has a high interest rate spread but low NPLs, while Egypt has low NPLs and a low interest rate spread. It is not possible to match the two moments in both countries without allowing for both limited commitment and costly state verification. Our bottom-line contribution is to allow and quantify the impact of different types of constraints on GDP, TFP, and inequality in an economy where all these frictions are potentially present.

Appendix

A Proofs

A.1 Proof of Proposition 1

For any level of capital, the optimal labor employed by entrepreneurs is given by the first order condition of problem (3.4),

$$l = [\frac{z(1-\alpha)(1-\nu)}{w}]^{\frac{1}{\alpha(1-\nu)+\nu}} k^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}. \tag{A.1}$$

Plugging l into (3.4), entrepreneurs solve

$$\begin{aligned} \pi^S(b,z) = \max_k \quad & (1-p)[(\tfrac{(1-\nu)(1-\alpha)}{w} z)^{\frac{1}{\alpha(1-\nu)+\nu}} w^{\frac{v+\alpha(1-\nu)}{(1-\nu)(1-\alpha)}} k^{\frac{\alpha(1-\nu)}{\nu+\alpha(1-\nu)}} - \delta k + k] \\ & + p\eta(1-\delta)k + (1+r^d)(b-k), \\ \text{subject to} \quad & k \;\; \leq \;\; b. \end{aligned}$$

Solving this problem without imposing the wealth constraint, the unconstrained capital demand is

$$\tilde{k}^S(z) = [\frac{1-p}{r^d + (1-p)\delta - p\eta(1-\delta) + p} \frac{\alpha w}{1-\alpha}]^{\frac{\alpha(1-\nu)+\nu}{\nu}} (\frac{(1-\nu)(1-\alpha)}{w} z)^{\frac{1}{\nu}}. \tag{A.2}$$

Since end-of-period wealth $\pi^S(b,z)$ is increasing in k for $k \leq \tilde{k}^S(z)$, the optimal capital for the constrained problem is

$$k^*(b,z) = \min\{b, \tilde{k}^S(z)\}. \tag{A.3}$$

A.2 Proof of Lemma 1

First we compute the default boundary $\tilde{\lambda}$. For entrepreneurs with no default risk, the recovered capital when production fails plus the amount of collateral (including interest earnings) should be higher than the face value of the loan. Therefore

$$\eta(1-\delta)k + (1+r^d)(b-\psi) \geq \Omega. \tag{A.4}$$

When condition (A.4) is satisfied, the zero profit condition (3.7) implies that $\Omega = (1+r^d)k$. Substituting this into (A.4), we obtain

$$\eta(1-\delta)k + (1+r^d)(b-\psi) \geq (1+r^d)k. \tag{A.5}$$

Using the definition of leverage ratio (3.6), (A.5) can be written as

$$\tilde{\lambda} = \frac{k}{b-\psi} \leq \frac{1+r^d}{1+r^d-\eta(1-\delta)}. \tag{A.6}$$

Hence, the default boundary is $\overline{\lambda} = \dfrac{1+r^d}{1+r^d-\eta(1-\delta)}$. Note that limited commitment imposes the constraint, $\tilde{\lambda} \leq \lambda$. To obtain a positive default rate for the model economy, we require $\dfrac{1+r^d}{1+r^d-\eta(1-\delta)} < \lambda$. This determines the range of endogenous interest rates $r^d > \dfrac{\eta(1-\delta)\lambda}{\lambda-1} - 1$.

Next, we compute the lending rate for entrepreneurs with leverage ratio $\tilde{\lambda}$.

If $\tilde{\lambda} \leq \overline{\lambda}$, entrepreneurs do not default. As stated above, the lending rate is equal to the deposit rate, $r^l = r^d$.

If $\tilde{\lambda} > \overline{\lambda}$, entrepreneurs default when production fails and condition (A.4) is violated. The zero profit condition (3.7) implies that the face value of the loan is

$$\Omega = \frac{(1+r^d)k + p\chi k - p\eta(1-\delta)k - p(1+r^d)(b-\psi)}{1-p}. \tag{A.7}$$

The lending rate defined by (3) is

$$r^l = \frac{1+r^d+p\chi - p\eta(1-\delta) - p(1+r^d)/\tilde{\lambda}}{1-p} - 1. \tag{A.8}$$

Note that the lending rate is discontinuous at $\overline{\lambda}$, $\lim_{\tilde{\lambda}\to\overline{\lambda}_+} r^l = r^d + \frac{p\chi}{1-p} \neq \lim_{\tilde{\lambda}\to\overline{\lambda}_-} = r^d$. This is due to the discontinuity of the optimal verification strategy as described in Proposition 2.

A.3 Proof of Lemma 2

For entrepreneurs in the low-leverage region ($\tilde{\lambda} \leq \overline{\lambda}$), default never happens, and entrepreneurs pay the interest rate $r^l = r^d$ regardless of whether production fails or not. Hence, the cost of capital is $R = r^d$.

For entrepreneurs in the high-leverage region ($\tilde{\lambda} > \overline{\lambda}$), when production succeeds, entrepreneurs pay the face value of the loan, Ω; when production fails, entrepreneurs default and pay $\eta(1-\delta)k + (1+r^d)(b-\psi)$. The cost of capital is equal to the expected amount of payment divided by the total amount of borrowing,

$$R = \frac{(1-p)\Omega + p[\eta(1-\delta)k + (1+r^d)(b-\psi)]}{k} - 1. \tag{A.9}$$

Substituting the zero profit condition (3.7) into (A.9), we obtain $R = r^d + p\chi$.

A.4 Proof of Proposition 3

In the credit regime, entrepreneurs solve problem (3.8) subject to the zero profit condition (3.7). This problem is non-convex because in the high-leverage region, banks monitor when production fails, increasing the cost of capital.

We solve the problem facing entrepreneurs of type (b, z) by converting problem (3.8) into two convex sub-problems: in one problem, entrepreneurs do not default, and leverage ratios are restricted by $\tilde{\lambda} \leq \overline{\lambda}$. In the other problem, entrepreneurs default when production fails, and leverage ratios are restricted by $\tilde{\lambda} \leq \lambda$.

The wealth function for each sub-problem is convex. The highest end-of-period wealth that can be obtained by entrepreneurs is the upper envelope of the two wealth functions, which is non-convex.

In the following, we first characterize the solution to each sub-problem and then provide the solution to the original problem.

The first sub-problem

In this sub-problem, entrepreneurs do not default. As shown in the proof of Lemma 1, $\Omega = (1+r^d)k$. Problem (3.8) can be written as

$$\begin{aligned} \pi^l(b,z) = \max_{k,l} \quad & 1-p)(z(k^\alpha l^{1-\alpha})^{1-\nu} - wl - \delta k - r^d k) + p\eta(1-\delta)k + (1+r^d)(b-\psi) - p(1+r^d)k \\ \text{subject to} \quad & k \quad \leq \quad \bar{\lambda}(b-\psi). \end{aligned}$$

By applying a similar analysis used in the proof of Proposition 1, we obtain the unconstrained level of capital

$$\tilde{k}^l(z) = [\frac{1-p}{r^d + (1-p)\delta - p\eta(1-\delta) + p} \frac{\alpha w}{1-\alpha}]^{\frac{\alpha(1-\nu)+\nu}{\nu}} (\frac{(1-\nu)(1-\alpha)}{w} z)^{\frac{1}{\nu}}.$$

The optimal amount of capital is

$$k^*(b,z) = \min(\bar{\lambda}(b-\psi), \tilde{k}^l(z)),$$

and the optimal amount of labor is

$$l^*(b,z) = [\frac{z(1-\alpha)(1-\nu)}{w}]^{\frac{1}{\alpha(1-\nu)+\nu}} k^*(b,z)^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}.$$

The wealth function is

$$\pi^l(b,z) = (1-p)(z(k^{*\alpha} l^{*1-\alpha})^{1-\nu} - wl^* - \delta k^* - r^d k^*) + p\eta(1-\delta)k^* + (1+r^d)(b-\psi) - p(1+r^d)k^*. \quad \text{(A.10)}$$

The second sub-problem

In this sub-problem, entrepreneurs default when production fails. As shown in the proof of Lemma 1,

$$\Omega = \frac{(1+r^d)k + p\chi k - p\eta(1-\delta)k - p(1+r^d)(b-\psi)}{1-p}.$$

Substituting this into (3.8), entrepreneurs solve

$$\begin{aligned} \pi^h(b,z) = \max_{k,l} \quad & (1-p)[z(k^\alpha l^{1-\alpha})^{1-\nu} - wl + (1-\delta)k + (1+r^d)(b-\psi)] \\ & -[(1+r^d)k + p\chi k - p\eta(1-\delta)k - p(1+r^d)(b-\psi)] \\ \text{subject to} \quad & k \;\leq\; \lambda(b-\psi). \end{aligned}$$

Similarly, we obtain the unconstrained level of capital

$$\tilde{k}^h(z) = [\frac{1-p}{r^d + p\chi + (1-p)\delta - p\eta(1-\delta) + p}\frac{\alpha w}{1-\alpha}]^{\frac{\alpha(1-\nu)+\nu}{\nu}}(\frac{(1-\nu)(1-\alpha)}{w}z)^{\frac{1}{\nu}}.$$

The optimal amount of capital is

$$k^*(b,z) = \min(\lambda(b-\psi), \tilde{k}^h(z)),$$

and the optimal amount of labor is

$$l^*(b,z) = [\frac{z(1-\alpha)(1-\nu)}{w}]^{\frac{1}{\alpha(1-\nu)+\nu}} k^*(b,z)^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}.$$

The wealth function is

$$\begin{aligned} \pi^h(b,z) = \quad & (1-p)[z(k^{*\alpha} l^{*1-\alpha})^{1-\nu} - wl^* + (1-\delta)k^* + (1+r^d)(b-\psi)] \\ & -[(1+r^d)k^* + p\chi k^* - p\eta(1-\delta)k^* - p(1+r^d)(b-\psi)]. \end{aligned} \tag{A.11}$$

The solution to the original problem is the upper envelop of the two sub-problems:

$$\pi^C(b,z) = \max\{\pi^l(b,z), \pi^h(b,z)\}.$$

To provide some intuitions on the optimal choice of the leverage ratio, consider two extreme cases:

(i) As $b \to \psi$, $\pi^h(b,z) \geq \pi^l(b,z)$. This is because the production function satisfies the inada condition. The marginal return is high when b is small and equality holds only when $b = \psi$.

(ii) As $b \to \infty$, $\pi^h(b,z) < \pi^l(b,z)$. This is because when entrepreneurs have sufficient wealth to operate businesses at the unconstrained scale, there is no reason to borrow from banks, which potentially increases the cost of capital due to monitoring.

Since both $\pi^l(b,z)$ and $\pi^h(b,z)$ are concave and increasing in b, there exists a unique intersection of the two curves, which defines the threshold level of wealth $\bar{b}(z)$. When b is below $\bar{b}(z)$, $\pi^l(b,z) < \pi^l(b,z)$, and the entrepreneur chooses high-leverage. Otherwise, the entrepreneur chooses low-leverage.

A.5 Proof of Proposition 4

Consider entrepreneurs with talent z, whose leverage ratios are denoted by $\tilde{\lambda}$. Note that $\bar{b}(z)$ is the threshold of wealth at which entrepreneurs are indifferent between taking low leverage and high leverage, i.e. $\pi^l(\bar{b}(z),z) = \pi^h(\bar{b}(z),z)$. Therefore, entrepreneurs with $b \in [\bar{b}(z), \bar{b}(z)+\epsilon)$ (with ϵ very small), are always hitting the borrowing constraint defined in the low-leverage region (i.e. $\tilde{\lambda} = \overline{\lambda}$).[31] Hence, the optimal amount of capital is $k_l^* = \overline{\lambda}(\bar{b}(z)-\psi)$ when $b = \bar{b}(z)$. Let l_l^* be the corresponding optimal amount of labor. The wealth function is

$$\pi^l(\bar{b}(z),z) = (1-p)(z((k_l^*)^\alpha(l_l^*)^{1-\alpha})^{1-\nu} - wl_l^* - \delta k_l^* - r^d k_l^*) + p\eta(1-\delta)k_l^* + (1+r^d)(\bar{b}(z)-\psi) - p(1+r^d)k_l^*.$$

In the high-leverage region, entrepreneurs with $b \in (\bar{b}(z)-\epsilon, \bar{b}(z))$ may or may not hit the borrowing constraint λ. Therefore, the optimal amount of capital is $k_h^* = \min(\lambda(b-\psi), \tilde{k}^d(z))$.

Let l_h^* be the corresponding optimal amount of labor. The wealth function is

$$\begin{aligned}\pi^h(\bar{b}(z),z) = \lim_{b\to\bar{b}(z)_-}\pi^h(b,z) &= (1-p)[z((k_h^*)^\alpha(l_h^*)^{1-\alpha})^{1-\nu} - wl_h^* + (1-\delta)k_h^* + (1+r^d)(\bar{b}(z)-\psi)] \\ &\quad -[(1+r^d)k_h^* + p\chi k_h^* - p\eta(1-\delta)k_h^* - p(1+r^d)(\bar{b}(z)-\psi)].\end{aligned}$$

Since $\pi^l(\bar{b}(z),z) = \pi^h(\bar{b}(z),z)$, $\bar{b}(z)$ is characterized implicitly by the following equation:

$$p\chi k_h^* = (1-p)[z[((k_h^*)^\alpha(l_h^*)^{1-\alpha})^{1-\nu} - ((k_l^*)^\alpha(l_l^*)^{1-\alpha})^{1-\nu}] - w(l_h^*-l_l^*)] + [p\eta(1-\delta)-p-(1-p)\delta-r^d](k_h^*-k_l^*).$$

Substituting l_l^* and l_h^*, we get

$$p\chi k_h^* = Ez^{\frac{1}{\alpha(1-\nu)+\nu}}((k_h^*)^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} - (k_l^*)^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}) + F(k_h^*-k_l^*), \tag{A.12}$$

where $E = (1-p)w\frac{\alpha(1-\nu)+\nu}{(1-\alpha)(1-\nu)}[\frac{(1-\alpha)(1-\nu)}{w}]^{\frac{1}{\alpha(1-\nu)+\nu}} > 0$, and $F = p\eta(1-\delta)-p-(1-p)\delta-r^d < 0$.

[31] The logic is as follows: a non-binding borrowing constraint implies that entrepreneurs can achieve the unconstrained level of capital $\tilde{k}^l(z)$ when $b \in [\bar{b}(z), \bar{b}(z)+\epsilon)$. Therefore, when $b \in (\bar{b}(z)-\epsilon, \bar{b}(z))$, entrepreneurs should also be able to achieve the unconstrained level of capital in the high-leverage region, $\tilde{k}^h(z)$, by borrowing only up to the leverage ratio $\overline{\lambda}$, since $\tilde{k}^h(z) < \tilde{k}^l(z)$. This is in contradiction to Proposition 3, which requires $\tilde{\lambda} > \overline{\lambda}$ when $b \in (\bar{b}(z)-\epsilon, \bar{b}(z))$.

To show that $\bar{b}(z)$ is increasing in z, we consider the following two cases.

Case 1

The borrowing constraint is binding for $b \in (\bar{b}(z)-\epsilon, \bar{b}(z))$, $k_h^* = \lambda(b-\psi)$.

Substituting k_l^* and k_h^* into equation (A.12), we obtain

$$[p\chi\lambda - F(\lambda-\bar{\lambda})](\bar{b}(z)-\psi) = Ez^{\frac{1}{\alpha(1-\nu)+\nu}}(\lambda^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} - \bar{\lambda}^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}})(\bar{b}(z)-\psi)^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}.$$

Take the first derivative with respect to z,

$$\bar{b}(z)' = \frac{\bar{b}(z)-\psi}{\nu z} > 0.$$

Case 2

The borrowing constraint is not binding for $b \in (\bar{b}(z)-\epsilon, \bar{b}(z))$.

In this case, $k_h^* = \tilde{k}^h(z) = [\frac{1-p}{r^d + p\chi + (1-p)\delta - p\eta(1-\delta) + p}\frac{\alpha w}{1-\alpha}]^{\frac{\alpha(1-\nu)+\nu}{\nu}}(\frac{(1-\nu)(1-\alpha)}{w}z)^{\frac{1}{\nu}} = Gz^{\frac{1}{\nu}}$, where $G = [\frac{1-p}{r^d + p\chi + (1-p)\delta - p\eta(1-\delta) + p}\frac{\alpha w}{1-\alpha}]^{\frac{\alpha(1-\nu)+\nu}{\nu}}(\frac{(1-\nu)(1-\alpha)}{w})^{\frac{1}{\nu}} > 0$.

Substituting k_l^* and k_h^* into equation (A.12), we obtain

$$Ez^{\frac{1}{\alpha(1-\nu)+\nu}}\bar{\lambda}^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}(\bar{b}(z)-\psi)^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} = [EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} + (F-p\chi)G]z^{\frac{1}{\nu}} - F\bar{\lambda}(\bar{b}(z)-\psi).$$

Take the first derivative with respect to z,

$$\bar{b}(z)' = \frac{\frac{[EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}+(F-p\chi)G]\frac{1}{\nu}z^{\frac{1}{\nu}}}{[EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}+(F-p\chi)G]z^{\frac{1}{\nu}}-F\bar{\lambda}(\bar{b}(z)-\psi)}\frac{1}{z} - \frac{1}{\alpha(1-\nu)+\nu}\frac{1}{z}}{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}\frac{1}{\bar{b}(z)-\psi} + \frac{F\bar{\lambda}}{[EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}}+(F-p\chi)G]z^{\frac{1}{\nu}}-F\bar{\lambda}(\bar{b}(z)-\psi)}}. \tag{A.13}$$

Next we show that both the numerator and denominator are smaller than 0, so that $\bar{b}(z)' > 0$.

Numerator:

$$F<0 \Rightarrow \frac{[EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} + (F-p\chi)G]\frac{1}{\nu}z^{\frac{1}{\nu}}}{[EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} + (F-p\chi)G]z^{\frac{1}{\nu}} - F\bar{\lambda}(\bar{b}(z)-\psi)}\frac{1}{z} - \frac{1}{\alpha(1-\nu)+\nu}\frac{1}{z} < \frac{1}{z} - \frac{1}{\alpha(1-\nu)+\nu}\frac{1}{z} < 0.$$

Denominator:

$$
\begin{aligned}
& [EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} + (F - p\chi)G]z^{\frac{1}{\nu}} \\
= & -(\frac{1}{r^d + p\chi + (1-p)\delta - p\eta(1-\delta) + p})^{\frac{\alpha(1-\nu)}{\nu}}[\frac{(1-\alpha)(1-\nu)}{w}]^{\frac{1}{\nu}}[\frac{(1-p)\alpha w}{1-\alpha}]^{\frac{\alpha(1-\nu)}{\nu}}(1-p)\frac{\alpha w}{1-\alpha}\frac{\nu}{(1-\nu)} \\
< & \ 0.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& \frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}\frac{1}{\bar{b}(z)-\psi} + \frac{F\bar{\lambda}}{[EG^{\frac{\alpha(1-\nu)}{\alpha(1-\nu)+\nu}} + (F - p\chi)G]z^{\frac{1}{\nu}} - F\bar{\lambda}(\bar{b}(z)-\psi)} \\
< & \ \frac{1}{\bar{b}(z)-\psi} + \frac{F\bar{\lambda}}{-F\bar{\lambda}(\bar{b}(z)-\psi)} \\
= & \ 0.
\end{aligned}
$$

In conclusion, $\bar{b}(z)' < 0$.

A.6 Proof of Lemma 3

When $\nu = 0$, according to equations (A.10) and (A.11), the wealth functions in the low-leverage and high-leverage regions are

$$
\begin{aligned}
\pi^l(b,z) &= (1-p)[\frac{\alpha}{1-\alpha}w[\frac{z(1-\alpha)}{w}]^{\frac{1}{\alpha}} + (1-\delta)]k \\
& \quad -(1+r^d)k + p\eta(1-\delta)k + (1+r^d)(b-\psi), \qquad (A.14) \\
\pi^h(b,z) &= (1-p)[\frac{\alpha}{1-\alpha}w[\frac{z(1-\alpha)}{w}]^{\frac{1}{\alpha}} + (1-\delta)]k \\
& \quad -(1+r^d)k + p\eta(1-\delta)k - p\chi k + (1+r^d)(b-\psi). \qquad (A.15)
\end{aligned}
$$

Substituting $k = \overline{\lambda}(b-\psi)$ and $k = \lambda(b-\psi)$ into equations (A.14) and (A.15), respectively. We obtain

$$
\begin{aligned}
\pi^h(b,z) - \pi^l(b,z) &= [(1-p)(\frac{\alpha}{1-\alpha}w(\frac{z(1-\alpha)}{w})^{\frac{1-\alpha}{\alpha}} + (1-\delta)) - (1+r^d) + p\eta(1-\delta)](\lambda - \overline{\lambda})(b-\psi) \\
& \quad -p\chi\lambda(b-\psi).
\end{aligned}
$$

$\pi^h(b,z) - \pi^l(b,z) > 0$ implies that

$$
z > \overline{z} = \frac{w}{1-\alpha}[(\frac{p\chi\lambda}{(\lambda-\overline{\lambda})} + (1+r^d) + p\eta(1-\delta) - (1-p)(1-\delta))\frac{1-\alpha}{\alpha w(1-p)}]^{\alpha}. \qquad (A.16)
$$

Therefore, talented agents choose the maximum leverage ratio λ.

A.7 Proof of Theorem 1

We focus on the case with $z > \overline{z}$. All proofs apply naturally to the case with $z \leq \overline{z}$.

When $\nu = 0$, using the results of Proposition 1, we obtain the wealth function in the savings regime as

$$\pi^S(b,z) = (1-p)[(\frac{1-\alpha}{w}z)^{\frac{1}{\alpha}}w\frac{\alpha}{1-\alpha} + (1-\delta)]b + p\eta(1-\delta)b. \tag{A.17}$$

Following Lemma 3, the wealth function in the credit regime is

$$\begin{aligned}\pi^C(b,z) &= (1-p)[(\frac{1-\alpha}{w}z)^{\frac{1}{\alpha}}w\frac{\alpha}{1-\alpha} + (1-\delta)]\lambda(b-\psi) \\ &\quad -[1+r^d+p\chi-p\eta(1-\delta)]\lambda(b-\psi) + (1+r^d)(b-\psi).\end{aligned}$$

Let $\pi^C(b,z) > \pi^S(b,z)$, we obtain

$$b > \underline{b} = \frac{(1-p)\lambda[(\frac{1-\alpha}{w}z)^{\frac{1}{\alpha}}w\frac{\alpha}{1-\alpha} + (1-\delta)] + 1 + r^d - \lambda(1+r^d+p\chi-p\eta(1-\delta))}{(\lambda-1)[(1-p)((\frac{1-\alpha}{w}z)^{\frac{1}{\alpha}}w\frac{\alpha}{1-\alpha} + (1-\delta)) - (1+r^d-p\eta(1-\delta))] - p\chi\lambda}\psi.$$

Let

$$\begin{aligned}m_1(z) &= (\frac{1-\alpha}{w}z)^{\frac{1}{\alpha}}w\frac{\alpha}{1-\alpha} + (1-\delta), \\ m_2 &= 1+r^d-p\eta(1-\delta).\end{aligned}$$

Then we can write $\overline{b}$ as

$$\underline{b} = \frac{\lambda((1-p)m_1(z)-m_2-p\chi)+1+r^d}{(\lambda-1)[(1-p)m_1(z)-m_2]-p\chi\lambda}\psi, \tag{A.18}$$

$$\pi^S(b,z) = [(1-p)m_1(z)-m_2]b + (1+r^d)b, \tag{A.19}$$

$$\pi^C(b,z) = [(1-p)m_1(z)-m_2-p\chi]\lambda(b-\psi) + (1+r^d)(b-\psi). \tag{A.20}$$

$z > \overline{z}$ implies that $\pi^h(b,z) - \pi^l(b,z) > 0$. Substituting $\pi^l(b,z)$ and $\pi^h(b,z)$ with equation (A.14) and (A.15), we obtain

$$[(1-p)m_1(z)-m_2]\lambda - p\chi\lambda > (1-p)m_1(z)\overline{\lambda} - m_2.$$

$\overline{\lambda} = \dfrac{1+r^d}{1+r^d-\eta(1-\delta)} > 1$ implies that

$$[(1-p)m_1(z)-m_2]\lambda - p\chi\lambda > (1-p)m_1(z) - m_2.$$

Thus,

$$(\lambda - 1)[(1-p)m_1(z) - m_2] > p\chi\lambda.$$

Therefore, $\underline{b}$ is well defined and $\underline{b} > 0$.

Take the first derivative,

$$\frac{\partial \underline{b}}{\partial \psi} = \frac{(1-p)m_1(z) + 1 + r^d - \lambda(m_2 + p\chi)}{(\lambda-1)[(1-p)m_1(z) - m_2] - p\chi\lambda} > 0, \tag{A.21}$$

$$\frac{\partial \underline{b}}{\partial \lambda} = -\frac{[(1-p)m_1(z) - m_2 - p\chi][1 + r^d + (1-p)m_1(z) - m_2]}{[(\lambda-1)((1-p)m_1(z) - m_2) - p\chi\lambda]^2}\psi < 0, \tag{A.22}$$

$$\frac{\partial \underline{b}}{\partial \chi} = \frac{(1-p)m_1(z) - m_2 + 1 + r^d}{[(\lambda-1)((1-p)m_1(z) - m_2) - p\chi\lambda]^2}\lambda p\psi > 0. \tag{A.23}$$

The second derivatives are:

$$\frac{\partial^2 \underline{b}}{\partial \chi \partial \lambda} = \frac{\partial^2 \underline{b}}{\partial \lambda \partial \chi} = -\frac{[1 + r^d + (1-p)m_1(z) - m_2][(\lambda+1)((1-p)m_1(z) - m_2) - \lambda p\chi]}{[(\lambda-1)((1-p)m_1(z) - m_2) - p\chi\lambda]^3}p\psi < 0, \tag{A.24}$$

$$\frac{\partial^2 \underline{b}}{\partial \lambda \partial \psi} = \frac{\partial^2 \underline{b}}{\partial \psi \partial \lambda} = -\frac{[(1-p)m_1(z) - m_2 - p\chi][1 + r^d + (1-p)m_1(z) - m_2]}{[(\lambda-1)((1-p)m_1(z) - m_2) - p\chi\lambda]^2} < 0, \tag{A.25}$$

$$\frac{\partial^2 \underline{b}}{\partial \chi \partial \psi} = \frac{\partial^2 \underline{b}}{\partial \psi \partial \chi} = \frac{(1-p)m_1(z) - m_2 + 1 + r^d}{[(\lambda-1)((1-p)m_1(z) - m_2) - p\chi\lambda]^2}\lambda p > 0. \tag{A.26}$$

A.8 Proof of Theorem 2

We focus on the case with $z > \overline{z}$. All proofs apply naturally to the case with $z \leq \overline{z}$.

When $\nu = 0$ and $z > \overline{z}$, the optimal labor demand is

$$l^*(b, z) = [\frac{z(1-\alpha)}{w}]^{1/\alpha}\lambda(b - \psi),$$

and

$$y^C = \pi^C + r^d\lambda(b - \psi) + (1-p)w[\frac{z(1-\alpha)}{w}]^{1/\alpha}\lambda(b-\psi) - (1 + r^d)b. \tag{A.27}$$

Hence,

$$\frac{\partial y^C}{\partial \psi} = -\lambda[(1-p)m_1(z) - m_2 - p\chi] - (1 + r^d) - r^d\lambda - (1-p)w[\frac{z(1-\alpha)}{w}]^{1/\alpha}\lambda < 0, \tag{A.28}$$

$$\frac{\partial y^C}{\partial \lambda} = [(1-p)m_1(z) - m_2 - p\chi + r^d](b - \psi) + (1-p)w[\frac{z(1-\alpha)}{w}]^{1/\alpha}(b-\psi) > 0, \tag{A.29}$$

$$\frac{\partial y^C}{\partial \chi} = -p\lambda(b - \psi) < 0. \tag{A.30}$$

$$\frac{\partial^2 \pi^C}{\partial \psi \partial \lambda} = \frac{\partial^2 \pi^C}{\partial \lambda \partial \psi} = -[(1-p)m_1(z) - m_2 - p\chi + r^d] - (1-p)w[\frac{z(1-\alpha)}{w}]^{1/\alpha} < 0, \quad \text{(A.31)}$$
$$\frac{\partial^2 \pi^C}{\partial \psi \partial \chi} = \frac{\partial^2 \pi^C}{\partial \chi \partial \psi} = p\lambda > 0, \quad \text{(A.32)}$$
$$\frac{\partial^2 \pi^C}{\partial \lambda \partial \chi} = \frac{\partial^2 \pi^C}{\partial \chi \partial \lambda} = -p(b-\psi) < 0. \quad \text{(A.33)}$$

B A Model with Forward-Looking Agents

Our model is based on an overlapping generation framework with a constant savings rate (or bequest rate). Therefore, the model is unable to capture one important way of coping with financial frictions: self financing. Buera, Kaboski and Shin (2011) and Moll (2014) all emphasize that the effects of financial frictions are amplified if a self-financing channel is precluded. As a result, it is expected that in the presence of forward-looking agents and endogenous savings rates, the impact of financial inclusion on GDP could be smaller.

In this appendix, we extend the model with forward-looking agents to address this concern.

B.1 Model Setup

We modify the baseline model with endogenous savings rates. The main modeling ingredients are similar to the baseline model, thus we state them briefly and only highlight the difference.

There is a continuum of agents living indefinitely. Population is constant and there is no aggregate uncertainty. Agents are heterogeneous in terms of wealth b and talent z. Wealth evolves endogenously, which is determined by agents' forward-looking decisions. Productivity z follows an exogenous Markov process. With probability γ, agents retain their productivities in the previous period; with probability $1-\gamma$, agents draw new entrepreneurial productivities. The new draw is from a time-invariant Pareto distribution governed by parameter θ and is independent of agents' previous productivities.

Agents have preference at time t,

$$E_t \sum_{s=t}^{\infty} \beta^{s-t} \frac{c_s^{1-\sigma} - 1}{1-\sigma}, \quad \text{(B.1)}$$

where β is the time discount factor.

Agents can choose occupation, either to become workers or entrepreneurs. Each worker supplies one unit of labor inelasticly and earns the equilibrium wage. Entrepreneurs use capital and hire labor to produce goods. Capital depreciates at rate δ after use. Entrepreneurs have access to the following production technology,

$$f(z,k,l) = z(k^{\alpha}l^{1-\alpha})^{1-\nu}. \tag{B.2}$$

Production fails with probability p, in which case output is zero and agents are able to recover only a fraction $\eta < 1$ of installed capital net of depreciation at the end of the period. Again, we assume workers get paid only when production is successful. Therefore, each worker receive the wage with probability $1-p$.

All agents can make a deposit at banks so as to transfer wealth across periods. However, agents need to pay a fixed credit participation cost ψ to borrow from banks. Borrowing is subject to limited commitment and asymmetric information problems as stated in the baseline model. Similarly, we use parameter λ and χ to capture the tightness of the borrowing constraint and the bank monitoring cost. Consistent with the myopic-agent model, we assume that agents' credit participation status is only maintained for one period. Therefore, agents who obtained credit in period t, still have to pay ψ in period $t+1$ if they want to borrow.

To simplify the problem (and also to be consistent with our myopic-agent model), we assume that agents choose occupation, credit participation, capital and labor to maximize expected end-of-period wealth (or expected income). Then, agents choose consumption and savings to maximizes utility. Notice that this allows us to solve the problem in two separate steps. In the first step, we solve a static problem to obtain the optimal occupation choice, credit participation, capital and labor inputs conditional on the beginning-of-period wealth and talent. In the second step, we solve a dynamic problem to obtain optimal consumption and savings. Thus the only difference from the baseline myopic-agent model is that we are endogenizing the consumption and savings decisions instead of using a constant savings rate.

Since the first part of the problem is solved exactly in the same way as the baseline myopic-agent model, below we only formulate the endogenous consumption/savings decisions, while taking the occupation choice, credit participation, capital and labor inputs as given.

Let $V(b,z,t)$ be the value function for agents of type (b,z) at the beginning of period t. Let I_t^s be the income if production succeeds, and I_t^f be the income if production fails in period t.

Therefore, given the occupation choice, credit participation, capital and labor inputs solved in the first part of the problem, I_t^s and I_t^f can be expressed as

$$I_t^s = \begin{cases} w_t & \text{Workers,} \\ z_t(k_t^{\alpha}l_t^{1-\alpha})^{1-\nu} - w_t l_t + (1-\delta-r_t^d)k_t + r_t^d b_t & \text{Entrepreneurs, savings regime,} \\ z_t(k_t^{\alpha}l_t^{1-\alpha})^{1-\nu} - w_t l_t + (1-\delta)k_t - \Omega + r_t^d b_t - \psi(1+r_t^d) & \text{Entrepreneurs, credit regime,} \end{cases}$$

$$I_t^f = \begin{cases} 0 & \text{Workers.} \\ -k_t + \eta(1-\delta)k_t + r_t^d(b_t - k_t) & \text{Entrepreneurs, savings regime.} \\ \max(0, \eta(1-\delta)k_t + (1+r_t^d)(b_t - \psi) - \Omega) - b_t & \text{Entrepreneurs, credit regime.} \end{cases}$$

Then agents choose consumption and savings to maximize life-time utilities. Denote c_t^s/c_t^f be the consumption when production succeeds/fails. Taking I_t^s and I_t^f as given, the recursive formuation for the second part of the problem is

$$\begin{aligned} V(b_t, z_t, t) = \max_{c_t^s, c_t^f, b_{t+1}^s, b_{t+1}^f} \quad & (1-p)[\frac{(c_t^s)^{1-\sigma} - 1}{1-\sigma} + \beta E[V(b_{t+1}^s, z_{t+1}, t+1)|z_t]] \\ & + p[\frac{(c_t^f)^{1-\sigma} - 1}{1-\sigma} + \beta E[V(b_{t+1}^f, z_{t+1}, t+1)|z_t]], \\ \text{subject to} \quad & c_t^s + b_{t+1}^s = b_t b_t + I_t^s, \\ & c_t^f + b_{t+1}^f = b_t + I_t^f. \end{aligned}$$

B.2 Calibration and Simulation Results

In the model with forward-looking agents, two extra parameters are introduced, the time discount rate β and the risk-aversion parameter σ. Following the standard practice, we set $\beta = 0.96$ and $\sigma = 1.5$. Computation complexity increases tremendously for the model with forward-looking agents. For tractability, we do not match the employment distribution, instead we set $\theta = 4.15$ for all the six countries following Buera and Shin (2013), which is selected to match the U.S. employment distribution. We choose parameters λ, ψ, p, χ, and η to match the collateral to loan ratio, the percent of firms with credit, the NPL ratio, the interest rate spread, and the bank overhead costs to assets ratio (see Table B.1). To obtain consistent comparisons with the baseline model, we re-calibrate all the parameters in the baseline model for each country under the parameter restriction, $\theta = 4.15$ (see Table B.2).[32]

The simulation results shown in Table B.3 indicate that, in general, the impact of relaxing the borrowing constraint on GDP and TFP is smaller (except for Mozambique), which is consistent with the results of Buera, Kaboski and Shin (2011). But the difference is not as large partly because in our model production fails with probability p, in which case entrepreneurs' wealth is wiped out. This constrains the self-financing channel. However, there are other differences. Notably, a reduction in the intermediation cost has a much larger GDP and TFP boosting effect as compared to the baseline model.[33] One reason for this could be that entrepreneurs respond to a

[32]By comparing the baseline model's simulation results with θ being calibrated to match country-specific firm-size distribution and θ being fixed at 4.15, we can analyze the impact of parameter θ on macroeconomic variables. However, according to the simulation results presented in Table B.3, the impact of this parameter on the model-predicted changes in GDP, TFP and the Gini coefficient is not as clear cut, depending on other country-specific characteristics.

[33]This seems to be consistent with the results in Greenwood, Sanchez and Wang (2013), which capture a wide

Table B.1: Calibration of the model with forward-looking agents

	Uganda			Kenya			Mozambique		
Target Moments	Data	Model	Parameter	Data	Model	Parameter	Data	Model	Parameter
Collateral (% of loan)	173	173	$\lambda = 1.58$	120.8	120.8	$\lambda = 1.83$	92	92	$\lambda = 2.09$
Firms with credit (%)	17.2	17.9	$\psi = 0.07$	25.4	22.93	$\psi = 0.08$	14.2	14.3	$\psi = 0.21$
Non-perfor. loan (%)	2.3	3.7	$p = 0.15$	10.6	10.9	$p = 0.18$	3.1	3.7	$p = 0.14$
Interest rate spread	10.9	9.5	$\chi = 0.80$	8.5	9.2	$\chi = 0.20$	8.2	7.4	$\chi = 0.7$
Overhead costs/assets	6.9	6.6	$\eta = 0.37$	6.6	6.5	$\eta = 0.37$	7.4	7.3	$\eta = 0.54$
	Malaysia			The Philippines			Egypt		
Target Moments	Data	Model	Parameter	Data	Model	Parameter	Data	Model	Parameter
Collateral (% of loan)	64.6	64.6	$\lambda = 2.56$	238.4	238.4	$\lambda = 1.42$	85.5	85.5	$\lambda = 2.17$
Firms with credit (%)	60.4	62.2	$\psi = 0.08$	33.2	31.6	$\psi = 0.07$	17.4	14.6	$\psi = 0.08$
Non-perfor. loan (%)	8.5	7.6	$p = 0.15$	4.5	3.4	$p = 0.11$	19.3	16.6	$p = 0.25$
Interest rate spread	3.3	5.0	$\chi = 0.05$	4.3	8.3	$\chi = 0.10$	6.1	8.8	$\chi = 0.01$
Overhead costs/assets	1.5	1.5	$\eta = 0.37$	3.2	3.1	$\eta = 0.29$	1.5	1.4	$\eta = 0.44$

lower intermediation cost by saving more, which is not captured if the savings rate is constant.[34] The same most-binding-constraints that constrain GDP are identified for all six countries by the two models. In Uganda, Kenya, Mozambique, and the Philippines, GDP increases more if the borrowing constraint is relaxed, while in Malaysia and Egypt, GDP is most responsive to a lower participation cost. The prediction on the change in income inequality is broadly consistent with the baseline model for reducing the participation cost and relaxing the borrowing constraint. However, the change in Gini coefficient has a flipped sign when the intermediation cost is lowered.

Table B.2: Calibration of the baseline model with $\theta = 4.15$

	Uganda			Kenya			Mozambique		
Target Moments	Data	Model	Parameter	Data	Model	Parameter	Data	Model	Parameter
Savings (% of GDP)	8	8	$\omega = 0.08$	15.4	15.4	$\omega = 0.15$	7.1	7.1	$\omega = 0.08$
Collateral (% of loan)	173	173	$\lambda = 1.58$	120.8	120.8	$\lambda = 1.83$	92	92	$\lambda = 2.09$
Firms with credit (%)	17.2	17.3	$\psi = 0.04$	25.4	25.4	$\psi = 0.10$	14.2	14.2	$\psi = 0.08$
Non-perfor. loan (%)	2.3	2.6	$p = 0.15$	10.6	5.7	$p = 0.17$	3.1	5.7	$p = 0.14$
Interest rate spread	10.9	9.7	$\chi = 0.85$	8.5	11.5	$\chi = 0.61$	8.2	10.8	$\chi = 0.73$
Overhead costs/assets	6.9	6.8	$\eta = 0.37$	6.6	6.4	$\eta = 0.45$	7.4	7.5	$\eta = 0.54$
	Malaysia			The Philippines			Egypt		
Target Moments	Data	Model	Parameter	Data	Model	Parameter	Data	Model	Parameter
Savings (% of GDP)	39	39	$\omega = 0.39$	25.7	25.7	$\omega = 0.26$	24.5	24.5	$\omega = 0.25$
Collateral (% of loan)	64.6	64.6	$\lambda = 2.56$	238.4	238.4	$\lambda = 1.42$	85.5	85.5	$\lambda = 2.17$
Firms with credit (%)	60.4	60.5	$\psi = 0.24$	33.2	33.2	$\psi = 0.07$	17.4	17.3	$\psi = 0.25$
Non-perfor. loan (%)	8.5	7.6	$p = 0.12$	4.5	3.2	$p = 0.11$	19.3	15.7	$p = 0.28$
Interest rate spread	3.3	5.7	$\chi = 0.15$	4.3	5.6	$\chi = 0.62$	6.1	8.0	$\chi = 0.08$
Overhead costs/assets	1.5	1.5	$\eta = 0.37$	3.2	3.2	$\eta = 0.27$	1.5	1.5	$\eta = 0.44$

Note: In this calibration, we set $\theta = 4.15$ for all six countries to be consistent with the calibration of the model with forward-looking agents. We choose parameters λ, ψ, p, χ and η to match the collateral to loan ratio, the percent of firms with credit, the NPLs ratio, the interest rate spread, and the bank overhead costs to assets ratio.

We prefer to retain our baseline overlapping generations framework as our primary specification. First, the theorems and propositions are clear in the baseline model and would not have closed-form expressions in a more general forward-looking-agents model, due to the concavity of the value

range of GDP per worker across different countries by purely varying the intermediation cost.

[34]This is confirmed by the simulated response of the credit access ratio. In the model with forward-looking agents, agents save to pay the credit participation cost as the monitoring cost decreases, which increases the credit access ratio significantly.

function. Second, computational complexity increases tremendously in the model with forward-looking agents.[35] This precludes the possibility of matching the employment distribution in every country.[36]

Table B.3: Comparing the impact of financial inclusion generated by the baseline model and the model with forward-looking agent.

		Participation cost ψ			Borrowing constraint λ			Intermediation cost χ		
		GDP(%)	TFP(%)	Gini	GDP(%)	TFP(%)	Gini	GDP(%)	TFP(%)	Gini
Uganda	Baseline	5.77	5.67	-0.0210	17.94	10.41	-0.0034	0.74	0.42	0.0018
	($\theta = 4.15$)	4.66	5.50	-0.0231	21.68	11.93	0.0233	0.74	0.35	0.0011
	FL model	7.46	1.84	-0.0087	9.65	3.80	-0.0119	6.89	2.11	-0.0165
Kenya	Baseline	5.16	6.50	-0.0314	12.28	9.30	-0.0203	1.93	0.74	0.0082
	($\theta = 4.15$)	4.79	5.39	-0.0300	13.90	10.28	-0.0014	1.70	0.49	0.0040
	FL model	6.98	4.24	-0.0206	10.68	6.16	-0.0263	6.18	2.56	-0.015
Mozambique	Baseline	12.72	10.16	-0.0267	10.30	4.83	0.0217	0.88	0.32	0.0033
	($\theta = 4.15$)	8.11	9.90	-0.0376	10.91	6.73	0.0081	0.87	0.24	0.0032
	FL model	12.08	6.70	-0.0174	13.32	6.70	-0.0111	10.78	6.36	-0.0241
Malaysia	Baseline	8.44	10.94	-0.0696	4.52	2.85	0.0059	1.26	0.00	0.0013
	($\theta = 4.15$)	6.62	6.92	0.0091	4.00	3.73	-0.0013	1.45	0.20	0.0048
	FL model	3.10	2.15	-0.0128	2.88	1.68	-0.0082	1.25	0.11	-0.0045
The Philippines	Baseline	2.56	3.40	-0.0165	20.21	16.45	-0.0336	1.48	0.58	0.0033
	($\theta = 4.15$)	2.16	3.34	-0.0167	22.43	18.02	-0.0124	1.48	0.73	0.0018
	FL model	10.98	5.02	-0.0226	18.58	8.56	-0.0550	7.75	2.17	-0.0092
Egypt	Baseline	7.04	11.31	-0.0590	7.78	6.61	0.0026	0.69	0.02	0.0033
	($\theta = 4.15$)	7.63	7.34	-0.0216	7.48	5.83	0.0025	0.74	0.08	0.0026
	FL model	5.81	2.00	-0.0502	5.68	2.00	0.0053	1.50	0.16	0.0015

Note: In all cases, we consider financial inclusion that moves the country to world financial sector frontier for one of the three parameters.

[35] For a computer with i7-4700MQ CPU (2.40GHz), it takes 20 minutes to compute the steady state of the baseline model using matlab (2014a). However, for the model with forward-looking agents, it takes more than 24 hours when a certain accuracy level is required. To promote the computation speed, we code the value function iteration and wealth distribution iteration parts of the program in C++ (VS studio 2014) and use matlab to call these scripts. At the same time, we use a 20-core server to parallel the computation of heterogeneity. This reduces the computation time of the steady state to 40 minutes. Hence, more complicated dynamics are within reach, but require more hardware and coding, which limits their wide applicability.

[36] As noted before, the calibration for the forward-looking-agents model is done by selecting parameter θ from the literature, which is calibrated using the U.S. employment distribution, not the distribution of each country. Similar approaches are also used in several other quantitative papers with forward-looking heterogeneous agents (e.g. Buera, Kaboski and Shin, 2011; Buera and Shin, 2013; Greenwood, Sanchez and Wang, 2013).

References

Abraham, Arpad, and Nicola Pavoni. 2005. "The Efficient Allocation of Consumption under Moral Hazard and Hidden Access to the Credit Market." *Journal of the European Economic Association*, 3(2-3): 370–381.

Aghion, Philippe, and Patrick Bolton. 1997. "A Theory of Trickle-Down Growth and Development." *Review of Economic Studies*, 64(2): 151–72.

Ahlin, Christian, and RobertM. Townsend. 2007. "Using Repayment Data to Test Across Models of Joint Liability Lending." *Economic Journal*, 117(517): F11–F51.

Albuquerque, Rui, and Hugo A. Hopenhayn. 2004. "Optimal Lending Contracts and Firm Dynamics." *Review of Economic Studies*, 71(2): 285–315.

Alem, Mauro, and Robert M. Townsend. 2013. "An Evaluation of Financial Institutions: Impact on Consumption and Investment Using Panel Data and the Theory of Risk-Bearing." *Annals Issue of the Journal of Econometrics in Honor of Bill Barnett*, , (forthcoming).

Amaral, Pedro S., and Erwan Quintin. 2010. "Limited Enforcement, Financial Intermediation, And Economic Development: A Quantitative Assessment." *International Economic Review*, 51(3): 785–811.

Andreoni, James. 1989. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *Journal of Political Economy*, 97(6): 1447–58.

Assuncao, Juliano, Sergey Mityakov, and Robert M. Townsend. 2012. "Ownership Matters: the Geographical Dynamics of BAAC and CommercialBanks in Thailand." Working Paper.

Ayyagari, Meghana, Asli Demirgc-Kunt, and Vojislav Maksimovic. 2008. "How Well Do Institutional Theories Explain Firms' Perceptions of Property Rights?" *Review of Financial Studies*, 21(4): 1833–1871.

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

Banerjee, Abhijit V, and Andrew F. Newman. 2003. "Inequality, Growth and Trade Policy." MIT Working Paper.

Beck, Thorsten, Asli Demirg-Kunt, and Ross Levine. 2007. "Finance, Inequality and the Poor." *Journal of Economic Growth*, 12(1): 27–49.

Beck, Thorsten, Asli Demirg-Kunt, and Vojislav Maksimovic. 2005. "Financial and Legal Constraints to Growth: Does Firm Size Matter?" *Journal of Finance*, 60(1): 137–177.

Buera, Francisco J., and Yongseok Shin. 2013. "Financial Frictions and the Persistence of History: A Quantitative Exploration." *Journal of Political Economy*, 121(2): 221 – 272.

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

Buera, Francisco J., Joseph P. Kaboski, and Yongseok Shin. 2012. "The Macroeconomics of Microfinance." National Bureau of Economic Research, Inc NBER Working Papers 17905.

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

Cagetti, Marco, and Mariacristina De Nardi. 2006. "Entrepreneurship, Frictions, and Wealth." *Journal of Political Economy*, 114(5): 835–870.

Caselli, Francesco, and Nicola Gennaioli. 2013. "Dynastic Management." *Economic Inquiry*, 51(1): 971–996.

Castaneda, Ana, Javier Diaz-Gimenez, and Jose-Victor Rios-Rull. 2003. "Accounting for the U.S. Earnings and Wealth Inequality." *Journal of Political Economy*, 111(4): 818–857.

Clarke, Georg R. G., Lixin Colin Xu, and Heng fu Zou. 2006. "Finance and Income Inequality: What Do the Data Tell Us?" *Southern Economic Journal*, 72(3): 578–596.

Clementi, Gina Luca, and Hugo A Hopenhayn. 2006. "A Theory of Financing Constraints and Firm Dynamics." *The Quarterly Journal of Economics*, 121(1): 229–265.

Davies, James B. 1982. "The Relative Impact of Inheritance and Other Factors on Economic Inequality." *The Quarterly Journal of Economics*, 97(3): 471–98.

Doepke, Matthias, and Robert M. Townsend. 2006. "Dynamic Mechanism Design with Hidden Income and Hidden Actions." *Journal of Economic Theory*, 126(1): 235–285.

Gine, Xavier, and Robert M. Townsend. 2004. "Evaluation of financial liberalization: a general equilibrium model with constrained occupation choice." *Journal of Development Economics*, 74(2): 269–307.

Greenwood, Jeremy, and Boyan Jovanovic. 1990. "Financial Development, Growth, and the Distribution of Income." *Journal of Political Economy*, 98(5): 1076–1107.

Greenwood, Jeremy, Juan Sanchez, and Cheng Wang. 2013. "Quantifying the Impact of Financial Development on Economic Development." *Review of Economic Dynamics*, 16(1): 194–215.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *The Quarterly Journal of Economics*, 124(4): 1403–1448.

Huggett, Mark. 1996. "Wealth Distribution in Life-cycle Economies." *Journal of Monetary Economics*, 38(3): 469–494.

Jeong, Hyeok, and Robert M. Townsend. 2008. "Growth And Inequality: Model Evaluation Based On An Estimation-Calibration Strategy." *Macroeconomic Dynamics*, 12(S2): 231–284.

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

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

King, Robert G, and Ross Levine. 1993. "Finance and Growth: Schumpeter Might Be Right." *The Quarterly Journal of Economics*, 108(3): 717–37.

Kinnan, Cynthia. 2014. "Distinguishing Barriers to Insurance in Thai Villages." Northwestern University Working Papers 1276.

Levine, Ross. 2005. "Finance and Growth: Theory and Evidence." In *Handbook of Economic Growth*. Vol. 1 of *Handbook of Economic Growth*, , ed. Philippe Aghion and Steven Durlauf, Chapter 12, 865–934. Elsevier.

Lloyd-Ellis, Huw, and Dan Bernhardt. 2000. "Enterprise, Inequality and Economic Development." *Review of Economic Studies*, 67(1): 147–68.

Martin, Alberto, and Filippo Taddei. 2013. "International Capital Flows and Credit Market Imperfections: A Tale of Two Frictions." *Journal of International Economics*, 89(2): 441–452.

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

Moll, Benjamin. 2014. "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?" *American Economic Review*, , (forthcoming).

Moll, Benjamin, Robert M. Townsend, and Victor Zhorin. 2014. "Economic Development and the Equilibrium Interaction of Financial Frictions." Working Papers.

Nardi, Mariacristina De. 2004. "Wealth Inequality and Intergenerational Links." *Review of Economic Studies*, 71(3): 743–768.

Okuda, Hidenobu, and Fumiharu Mieno. 1999. "What Happened to Thai Commercial Banks in the Pre-Asian Crisis Period: Microeconomic Analysis of Thai Banking Industry." *Hitotsubashi Journal of Economics*, 40(2): 97–121.

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.

Townsend, Robert M. 1979. "Optimal Contracts and Competitive Markets with Costly State Verification." *Journal of Economic Theory*, 21(2): 265–293.

Townsend, Robert M., and Kenichi Ueda. 2006. "Financial Deepening, Inequality, and Growth: A Model-Based Quantitative Evaluation." *Review of Economic Studies*, 73(1): 251–293.

World Bank. 2014. "Global Financial Development Report: Financial Inclusion."