Financial frictions, financial market development, and macroeconomic development

Joseph P. Kaboski*
University of Notre Dame

October 6, 2021

Abstract

This paper reviews the state of knowledge on the impact of financial frictions and financial underdevelopment on firms. The focus is on their aggregate and distributional consequences for the macroeconomies of developing countries. It then reviews available data and data needs for future progress and proposes an agenda of important but unanswered questions for informing our understanding of growth and policy. Various questions involve ways to promote financial development itself, guide second-best policies in the face of financial frictions that enable macrodevelopment, and develop an inclusive financial system that benefits all.

*This paper was prepared as a Pathfinding Paper for the CEPR-led research programme on Structural Transformation and Economic Growth (STEG), funded by the United Kingdom’s Department for International Development (DFID) as part of the UK aid effort. Contact: University of Notre Dame, Department of Economics, 3039 Nanovic Hall, Notre Dame IN 46556, jkaboski@nd.edu.
1 Introduction

Financial underdevelopment is a common characteristic of poor countries. In 2017, the ratio of intermediated private credit to GDP was just 31 percent for low and lower middle income countries, as compared to 86 percent for high income countries. After taking into account the underdevelopment of stock and private bond markets, the differences in available external finance to firms are even starker.\footnote{The numbers are based on available data World Bank’s Financial Structure Database, where intermediated private credit is defined as private credit by deposit money banks and other financial institutions. If one adds in the stock market capitalization and private bond market capitalization (for countries with available data), the ratios average 38 percent for low and lower middle income countries, and 159 percent for high income countries.} Thus, firms may face stronger constraints to their investment, cash flow management, and growth because of financial underdevelopment and the financial frictions involved.

This paper reviews the state of knowledge on the impact of financial frictions and financial underdevelopment on firms. While the correlation between financial development and economic development is strong, there is longstanding debate about the strength of the causal relationship. The paper discusses recent work that more effectively identified causal channels, empirically and using quantitative theory. It then reviews available data, describes data needs for future progress, and proposes an agenda of important but unanswered questions for informing our understanding of growth and policy.

The idea that finance was important for economic development has a long history. Smith (1776) is perhaps most famous for theorizing about the importance of finance in enabling entrepreneurs. Still, there have been notable skeptics as well, including Robinson (1952) and Lucas (1988). There has always been the claim that in the relationship between financial and economic development causation goes from growth to finance. The first wave of empirical research evaluating this claim used cross-country regression techniques to try and tease out causation, such as King and Levine (1993a), who used timing assumptions to argue that financial development caused future growth, and Rajan and Zingales (1998), who examined whether financial development led to comparative advantage and differential growth in external finance dependent industries.

In recent decades this cross-country literature has been supplemented by empirical micro
evidence of financial frictions from developing countries. Empirically, small and large firms show evidence of being financially constrained. Microfinance and simple grants have been helpful in promoting investment and profitability in some cases. Nonetheless, even at the micro level, programs need to be well targeted and screening borrowers is critical.

The literature has also been supplemented by quantitative models of financially constrained entrepreneurs that give insights into the potential magnitudes of the impact of finance on economic development and the mechanisms by which entrepreneurs are impacted. In line with the empirics, the models show that the behavior of highly productive entrepreneurs with high growth potential is very different from that of subsistence or necessity entrepreneurs. Hence, a recurring policy lesson is the importance of targeting for interventions, and the likely importance of better screening mechanisms for financial development. Another policy lesson is that this environment is very dynamic. Finance can more easily adapt to ever changing needs of entrepreneurs, whereas one-time redistributions can quickly evaporate.

Financial development can also enhance structural transformation. Financial frictions disproportionately hurt industries which rely more heavily on external finance, e.g., large-scale production, heavy manufacturing, RD or marketing intensive sectors, at the expense of sectors that can be more easily self-financed (e.g., small-scale retail, subsistence farming). Firm growth itself can be a pull factor out of agriculture and small-scale services.

Moving forward there are some other policy lessons, policy questions, and methodological lessons of how to connect micro empirical evidence, including experimental evidence, with macro models in order to inform policy. Research needs to ask more aggregate policy questions and develop policy lessons, and the paper suggests several key avenues: identifying keys to financial development, determining second-best policies, and evaluating policies for financial inclusion when scaled to the aggregate level.

The focus of this paper is firms, which is only one aspect of the impacts of financial development. Finance can also be important for agriculture, where it can improve productivity by enabling investments in modern technologies. Additionally, intermediate usage can improve productivity and act as a push factor for structural transformation, and there may also be an important role that finance can play for consumer finance.
The rest of this paper is organized as follows. In Section 2, I discuss the literature on the quantitative importance of financial frictions and financial development for macroeconomies. Section 3 reviews the literature on microinterventions and how they relate to macropolicy. In Section 4, I discuss existing data and needed data to better evaluate these models empirically and enhance understanding of situations in developing countries. Section 5 discusses pressing questions for future research, while Section 6 concludes.

2 Quantifying the Macro Impact of Frictions

In theory, financial development can positively impact economic growth. Savings instruments can mobilize capital. Credit can channel capital toward its most productive uses facilitating both trades and investment. Intermediaries can pool information and resources for large investments while providing liquidity. Beyond banks, securities markets can allow for broader risk diversification and mediation. Both institutions can exert monitoring and corporate control, allowing for the effective separation of management and financial stakes. This can enable those without wealth access to productive capital, an equalizing force in theory. Yet, there remains the sense that finance can also be a parasite on the economy that creates real economic fragilities while benefiting only the wealthy.

2.1 Reduced Form Macro Empirical Analysis

Empirically, I have touched on the aggregate evidence of finance on development, including a high correlation of the size of the financial sector with average incomes in the cross section of countries and over time series of advanced economies. In addition, panel evidence shows that aggregate measures of financial intermediation such as liquid liabilities (M3) over GDP or non-financial private credit are leading indicators of aggregate growth (King and Levine, 1993a), with initial values predicting future growth of income, investment, capital and total factor productivity (TFP) growth. Noting that financial variables and decisions are, by nature, forward looking, Rajan and Zingales (1998) look for alternative evidence to Granger-like causation by examining whether industries that are highly dependent on external finance grow disproportionately faster in countries that are financially more developed.
Their measures included domestic credit and stock market capitalization relative to GDP, while they measured external dependence by industry using U.S. firm level data. They found strong evidence that industries that were more financially dependent had faster growth in value-added and the number of establishments between 1980 and 1990. However, the robustness of their results, especially to extended samples, has been criticized. Moreover, Do (2007) have argued that financial development itself is dependent on comparative advantage patterns. Jayaratne and Strahan (1996) apply a difference-in-difference strategy to a panel of U.S. states that experienced varying liberalizations of banking restrictions and showed that financial development that stemmed led to faster state income growth. 2

This more aggregate empirical work has studied overall financial development and its dimensions (e.g., stock markets, private credit, savings, securities markets, regulation, competition), but it has generally focused on either cross-country studies or advanced economies. Moreover, the focus of most of the growth theory is the role of finance for firms and entrepreneurs, but the broad empirical work can mix together other sizable elements of the financial system like consumer credit and mortgages. Although money is fungible, such studies on their own may be limited in how much insight they can offer on the nature of frictions that firms or potential entrepreneurs in developing countries face.3

2.2 Microevidence of Financial Frictions

A complementary literature has emerged emphasizing the micro financial frictions that many firms face in developing countries. One goal of the financial system is to redirect capital from those with low marginal returns to investing to those with high marginal returns through saving and borrowing, but this channel is limited when the wedge between interest rates of savings and credit is sizable. Nevertheless, a sizable gap between lending and deposit rates is a striking feature of financial systems, especially in developing countries. For instance, according to the International Financial Statistics, the average interest rate spread (lending

---

2 A recent review of this empirical literature is Levine (2018).
3 Several studies use careful quasi-experimental variation at a meso-level of aggregation (Burgess and Pande, 2005; Breza and Kinnan, 2018; Bustos et al., 2020), including in conjunction with structural models (Kaboski and Townsend, 2012), to look in more depth at household finance and its interplay with investment and firm growth, and to give evidence of microfrictions.
minus deposit rate) is approximately 0.7 percent in Japan, 3 percent in the United States, and 5 percent in Italy, but Banerjee (2003) shows spreads of 30 to 60 percent are common in developing countries. In Brazil, Cavalcanti et al. (2019) find that the standard deviation of spreads on formal loans to firms in Brazil, after controlling for past and future default, is 32 percentage points. The goal of equalizing the bang for buck on marginal capital invested is undermined by this variation in spreads.

Additional evidence of financial frictions comes from evidence on returns to capital. A prime example comes from experimental cash grants to entrepreneurs with returns ranging from 4-6% per month (in Sri Lanka) up to 33% per month (in Mexico) as reviewed by Buera et al. (2019). Moreover, these cash grants led not only to high returns but sizable increases in the capital stock. There is also evidence from the microfinance literature, where credit can have substantial impacts for some entrepreneurs (Banerjee et al., 2019, e.g.). All of this is evidence of either credit-constrained entrepreneurs — quantity constrained at reasonable interest rates — or entrepreneurs facing high marginal costs of borrowing. Here the experimental evidence is overwhelmingly on micro or small enterprises, however.

Firm size distributions and rates of entrepreneurship provide more indirect evidence of the prevalence of broad financial frictions. Average firm sizes (measured as the number of employees) are substantially lower in poorer countries. For many years, it was thought that there was a “missing middle” of the firm size distribution in poor countries, but there is little evidence for this given representative, detailed data (Hsieh et al., 2014). Firms in developing countries have different growth patterns as well, with the firms in poorer countries like Mexico and India growing much less on average than firms in the United States (Hsieh and Klenow, forthcoming). Of course, informal firms are much more common in poor countries, as is entrepreneurship in general (Poschke, 2013).

To the extent that differences in firm size distributions reflect distortions rather than differences in the labor, demand, or technology, these firm size distributions may be the result of firms facing financial frictions. While some small firms may have high marginal products of capital, on average large firms have higher average products of capital and labor (Hsieh et al., 2014). Lacking sufficient access to capital may limit growth of large firms. The lower demand for labor among highly productive firms can drive higher rates of self-
employment and necessitate entrepreneurship.

2.3 Aggregate Theory

Another limitation of the aggregate growth-finance literature is in understanding mechanisms sufficiently to extrapolate crisp policy lessons to developing countries. For example, Jayaratne and Strahan (1996) measured the growth impacts of the relaxation of restrictions on bank branching within U.S. states, but what does this teach us about developing countries? Will all bank branching deregulation lead to growth, or (more likely) does it depend on the environment and type of frictions firms face?

Addressing questions like these necessitates theory, and so I briefly review the theoretical literature on financial development and growth. Important early theoretical contributions include Evans and Jovanovic (1989); Greenwood and Jovanovic (1990); Boyd and Smith (1992); Bencivenga and Smith (1991); Banerjee and Newman (1993); Galor and Zeira (1993); King and Levine (1993); Aghion and Bolton (1997); Lloyd-Ellis and Bernhardt (2000); Ghatak and Jiang (2002) focused on mechanisms without reference to quantitative importance.

Quantitative models are useful for examining the potential gains to financial development, understanding mechanisms to inform policy, and for running counterfactual policy simulations. A literature of quantitative modeling has developed strongly over the past twenty years, including the pioneering work on the Thai growth experience by Townsend and coauthors (Gine and Townsend, 2004; Jeong and Townsend, 2007; Townsend and Ueda, 2006; Townsend, 2010b; Townsend and Ueda, 2011). This literature models the behavior of entrepreneurs facing financial frictions, focusing on the core growth mechanism (Buera et al., 2011; Greenwood et al., 2010, 2013; Cole et al., 2013; Midrigan and Xu, 2014; Moll, 2014). The models typically involve entrepreneurial entry decisions, dynamic savings decisions, heterogeneity in productivity, and uncertainty along with financial frictions. They can be augmented to allow for household volatility in labor income or employment opportunities, the lack of which can lead to “necessity entrepreneurs”. The prominence of such necessity entrepreneurs in developing countries have been well documented empirically (Paulson and Townsend, 2004; Poschke, 2013; ?; Adhvaryu et al., 2019; Karaivanov and Yindok, 2020).

Computing and solving these models can be challenging. These models are complicated
because prices are functions of the state, which is an entire distribution. Moreover, the time path of prices is important for forward looking behavior, so solving these models outside of stationary equilibria is even more challenging. Additionally, the presence of discrete decisions like occupational choice or technology upgrading can introduce kinks into value functions that create their own challenges. Nevertheless, improvements have been made in methods to approximate solutions, computational speed, and limiting the state space for dynamic decisions, all of which makes these more tractable (see, e.g., Achdou et al., n.d.).

As with its purely theoretical precedents, there has been disagreement about how important financial frictions are quantitatively. Nevertheless, several important lessons have emerged. First, the joint distribution of productivity and wealth drives the strength of quantitative results. When wealth and productivity are aligned, there is less of an efficiency rationale or need for the financial system to reallocate resources. Second, the persistence of productivity is an important determinant of this joint distribution. If productivity is highly persistent, wealth and productivity will more naturally align, whereas if productivity is highly transient, there constant reallocation of capital will be needed. Third, the ability to self-finance is a potentially strong way of overcoming credit constraints. This is easier to do, the larger the interest rate on saving. Of course, the ability to self-finance depends greatly on the persistence of productivity as well. Fourth, the ability for large firms to access finance is particularly important for aggregates. When these firms can access finance more easily, either because of accumulated wealth or borrowing limits that increase with productivity/income, financial frictions costs are lower.

A fifth important lesson is the importance of discrete decisions like fixed costs, entry costs, and one-time technology choices. For example, Buera et al. (2011) find that in a two-sector model calibrated to the United States, when there is a substantial fixed cost to enter the large-scale sector, the aggregate costs of financial frictions can amount to a factor of two difference between financial autarky and perfect credit markets. They find that entry distortion is one of the big costs, and that this cannot be observed by simply analyzing misallocation among the full set of existing firms as in, for example, Hsieh et al. (2008).

A caveat is if productivity is endogenous and wealth is important to finance productivity-improving investments. The endogenous impact of wealth on productivity has been examined empirically (Bell et al., 2018), theoretically (Gale, 1990), and quantitatively (Buera and Fattal-Jaef, 2014).
Moreover, lower entry and productivity in the large-scale sector can depress wages, leading to excess entry in the small-scale sector where financial frictions are not prohibitive. This is a theory for high levels of entrepreneurship in small-scale sectors (e.g., menial services), and can drive down the relative price of output in those sectors. Midrigan and Xu (2014) calibrate their model with a traditional and modern technology choice, to firm panel data in South Korea and Colombia, and find that productivity is extremely persistent among existing firms and entry costs and into a modern technology are the most important source in driving the potential costs of financial frictions.

Both studies find that indivisible one-time costs can be harder to finance than large fixed costs. While a series of fixed costs might be paid out of current profits, a one-time entry or technology cost of equivalent discounted value needs to be financed up front, and self financing is difficult because income may be much lower before entry or upgrading. Since heavy industrial and large-scale services (e.g., hospitals) tend to have large fixed costs, these models provide theory for how financial frictions can disproportionately impact large-scale sectors and structural transformation at the industry level, reducing productivity in these industries and pushing resources into smaller-scale sectors. Nevertheless, at smaller levels, nonconvexities and fixed costs can play important role even in other sectors like agriculture (Balboni et al., 2018; Kaboski et al., 2019), carpentry (Bassi et al., 2020), and petty trading (Banerjee et al., 2019).

Midrigan and Xu (2014)’s panel firm data allow them to examine firm dynamics to quantitatively discipline their model. Firm dynamics can potentially inform us about the strength and nature of frictions that firms face. In the models of Buera et al. (2011) and Midrigan and Xu (2014), firms all face a common interest rate, and the financial constraint takes the form of a quantity constraint. High productivity entrepreneurs who are active but highly constrained have a high marginal product of capital that exceeds the interest rate, and so they re-invest the majority of their earnings back into the business. They therefore accumulate assets, weakening the constraint, and grow quickly. Since Midrigan and Xu (2014) do not see large differences in the average growth rates of young and old firms, they determine that misallocation among existing firms cannot explain much productivity loss at all.
3 Integrating Micro Empirics and Macro Models for Policy Evaluation

We therefore have strong evidence and arguments from various methodologies for financial development leading to aggregate growth. Integration of these methodologies, especially micro-empirics with quantitative macro models, shows great promise in providing more insights into scaled and macro policy evaluation. I highlight two areas for integration in particular: using micro evidence to discipline macro models and then using macro models to extend policy analysis to the macroeconomy. As examples, I focus on the sources and types of financial frictions and the presence of poverty traps.

3.1 Sources and Types of Frictions

The type and underlying source of financial frictions can impact macro analysis, and microevidence can give insight into the relevant types of frictions. An important example where a data-based inference is not robust is the above-mentioned result on firm growth and misallocation of Midrigan and Xu (2014). Most models, like Buera et al. (2011), Midrigan and Xu (2014), and Moll (2014) model financial frictions as a reduced form capital borrowing limit, which depends on wealth and productivity. The two are designed to capture the assets and future income flow usually used to assess loans.

However, the model in Cavalcanti et al. (2019) introduces heterogeneous credit spreads as an additional friction to match their credit registry and firm panel data from Brazil. At a point in time, there is no difference in output or productivity of a firm who has less capital because of a quantity borrowing limit or a firm who has the same level of capital because they face a high interest rate. However, there is a dynamic difference between the two; the latter has a more difficult time accumulating assets, as its high returns on capital cannot be reinvested and accumulated because it goes to pay the high cost of capital. Thus, constrained firms have difficulty growing out of their constraint. The quantitative implications of spread frictions are therefore much stronger than equivalent collateral constraints. Cole et al. (2013)
find a similar result from a very different rationale. They use a model of uncertainty and costly state verification, but they endogenize the distribution of firm productivity and make finance essential for innovation. Financial frictions therefore slow productivity growth, and they find large impacts of financial frictions without entry or misallocation playing any role. In Cole et al. (2013), firm sizes are again smaller, and slow growth is no longer an indication of a lack of a financial constraint. Indeed, the quantitative impacts of financial frictions are substantially larger than the factor of two difference in the papers on self-financing.

Distinguishing between the sources of financial constraints can matter both quantitatively and for policy prescriptions. Beyond limited liability or enforcement, which can motivate the collateral constraints in the above models, moral hazard and adverse selection (and the monitoring costs they can impose) are other chief underlying sources of financial constraints.

Relatively less research delves into distinguishing the source of microfrictions, but we do have some microevidence on channels. Paulson et al. (2006), Karaivanov and Townsend (2014), and Moll et al. (2017) find that moral hazard is important in central and semi-urban areas, but that other regimes prevail in rural areas of Thailand. Working in rural and urban areas of South Africa, Karlan and Zinman (2009) find evidence for moral hazard as well, but they emphasize that men suffer from moral hazard, while adverse selection impacts women. Thus, the evidence is not simple or everywhere uniform, but moral hazard seems to play a role but it is unclear if these patterns hold for larger firms.

The source of friction can matter for quantitative impacts of financial frictions and potential policy remediation. As noted, limited liability constraints are wealth dependent constraints that can be grown out of over time, but as assets grow, borrowing grows as well. Hence, not all borrowers are constrained. Moral hazard can lead to low effort or fraud in reporting income to intermediaries. When effort is at the core of the moral hazard problem, those with sufficient wealth opt not to borrow (Aghion and Bolton, 1997), thereby avoiding agency costs. When moral hazard is key, financial frictions can therefore take the form of credit spreads for various reasons: monitoring costs, costly state verification, differential risks that lead to differential monitoring costs, and differential levels of collateral.6 It can also lead to dynamic incentives in contracting. When moral hazard constraints bind, credit bureaus

---

6Moral hazard can also lead to collateral based rationing as in Holmström and Tirole (1998).
may be effective, since they increase the penalties of low effort and default. In a world of adverse selection, however, credit bureaus can undermine financial markets by eliminating the long term benefits of financial institutions taking costly efforts (e.g., initial loans to those without credit records) of acquiring information and identifying good customers.\footnote{Adverse selection can lead to credit rationing as in Stiglitz and Weiss (1981) or overprovision of credit as in de Meza and Webb (1987).}

At the same time, in a world with a lot of variation in entrepreneurial productivity and profitability and adverse selection, screening can be critical, and so there is room to look for innovations in gathering information. The experimental work on microfinance of Banerjee et al. (2019) shows that existing microentrepreneurs are much more likely to be high productivity in India. The micro, experimental work of Hussam et al. (2018) uses cash grants and incentive experiments to show not only that heterogeneity is important and screening can increase the profitability of marginal capital injections, but they also show how to screen: mechanism design can be used to elicit helpful community information on individuals and such methods could be cost effective for lenders. A good question is whether such methods would scale appropriately to help foster financial development and inclusion.

3.2 Poverty Traps

3.2.1 Empirical Evidence

Another area in which microevidence and macrotheory can work together is in assessing the importance of poverty traps. Several studies give us evidence on poverty traps. First, there is long run experimental evidence of the impacts of cash grants. Second, there are experimental and time series studies of poverty traps themselves.

Earlier experimental evidence from asset grant interventions in Uganda showed positive impacts on assets, income, and consumption (Blattman et al., 2014). However, in harmony with the Buera et al. (2014) predictions, some more recent experimental evidence shows no \textit{long term} impacts on earnings and consumption (Blattman et al., 2018). Similarly, in Ethiopia, Blattman and Dercon (2018) gave cash grants to one treatment branch of unemployed youths, and industrial jobs to another branch. Five years out, the treatment and control branches were completely indistinguishable in terms of employment, income,
and health (Blattman et al., 2019). However, two other RCTs find sustained impacts from asset grants even many years out: grants to existing entrepreneurs in Sri Lanka (de Mel et al., 2012), and livestock grants to poor women in Bangladesh Bandiera et al. (2013). The former study may have better targeted poor yet productive entrepreneurs. The latter study attributes the impacts to an increase in labor supply. My point is to emphasize that the evidence is mixed and to stress that variations in underlying environments and populations may lie behind the disagreement. Careful modeling, with assumptions appropriate to both the environment and policy, are important.

Using long run observational data, Lybbert et al. (2004) and Santos and Barrett (2011) show that livestock dynamics of pastoralists in Ethiopia exhibit multiple equilibria. They also link informal credit targeting to the threshold values that define the multiple equilibria dynamics. Balboni et al. (2018) examines the above-mentioned livestock grant programs in Bangladesh. In that study, all households were given the same sized grant, and researchers examined how initial wealth determined trajectories. Focusing on asset dynamics, they show s-shaped returns consistent with poverty trap asset dynamics, and evidence of twin peaks in the asset distribution, consistent with multiple equilibria. Banerjee et al. (2019) look at the dynamics from a microcredit experiment in India and estimate a poverty trap with threshold levels of capital, but only for pre-existing, “gung ho” entrepreneurs. (Hence, the importance of targeting or screening is again pre-eminent.) Examining the same issue in Uganda – though not related to above-mentioned Ugandan asset grants – Kaboski et al. (2019) use a choice across different lotteries as an alternative way of combining intensive margin variation with randomization, and tying it to risk preference. When agents are near threshold values or areas of increasing returns, they can be locally risk-loving, willing to take chances to escape poverty traps. The authors offer a choice over cash lotteries to an underbanked population in Uganda and show that a substantial share of the population prefers a high-risk lottery even when it has a slightly lower expected payout.

This tells us that, in various settings, indivisibilities may be quite important for the poor and microenterprises, but it tells us little about the relevance for those potential entrepreneurs who would start the larger firms that drive the macroeconomy, especially as economies grow. Nor does it tell us whether such micro poverty traps lead to macro poverty
traps for entire economies.

### 3.2.2 Macro Theory

Macro theories study the latter. The idea of that whole macroeconomies could be stuck in poverty traps dates back at least to the mid-20th century “Big Push” theories of development, in which strong complementarities drove multiple equilibria. More narrowly, the idea that financial frictions, entrepreneurship decisions, and the distribution of wealth can lead to poverty traps was a common theme in the theoretical literature in the mid-1990s, (e.g., Banerjee and Newman, 1993; Galor and Zeira, 1993; Piketty, 1997; Ghatak and Jiang, 2002).

Poverty traps arise in these models at both the micro level and the macro level. The starting assumption is a non-convexity such that entrepreneurs require a minimum scale of capital to be profitable. Those with enough wealth to finance the minimum scale enter a virtuous cycle of entrepreneurship, income, and wealth accumulation, while the incomes of those without sufficient wealth are determined by low wages and interest rates, and their wealth decumulates. This is the micro-level poverty trap. Macro poverty traps arise because initial wealth distributions affect general equilibrium wages or interest rates, and ultimately impact aggregate dynamics. If most people lack the required initial assets to become entrepreneurs, for example, there will be few entrepreneurs. Consequently, the demand for labor and capital will be low and so will wages and interest rates, but entrepreneurial income will be high. These returns will perpetuate a wealth distribution in which few have the resources to become entrepreneurs. These models of poverty traps came with policy prescriptions as well: Since the equilibria were distribution dependent, one-time wealth redistributions could jolt the economy from the bad to the good equilibrium, improving aggregate surplus.

In thinking about macro poverty traps, the quantitative models have made an important contribution. Aghion and Bolton (1997) showed that equilibria could be unique if there were sufficient levels of savings and capital accumulation. In that case, there were still aggregate costs to financial frictions, but no aggregate poverty trap. More generally, the macro poverty traps arose because of the stylized nature of the models. For example, in Banerjee and Newman (1993) and Ghatak and Jiang (2002), there is no intensive margin in the demand for labor that would make equilibrium wages respond continuously rather than discretely.
In Piketty (1997) and Aghion and Bolton (1997) there is no intensive margin for capital which could respond to the interest rate. Savings behavior is modeled with “warm glow” motives, which precludes dynamic self-financing motivations to overcome constraints. These assumptions are relaxed in the more recent quantitative literature by adding in intensive margins for businesses and forward-looking behavior. Moreover, they incorporate additional features of realism such as high levels of heterogeneity in productivity, which we observe in the experimental literature, and churning, which is consistent with the long run dissipation of the impacts of asset grants. These features are also needed to map models to observational data on firm size distributions and firm dynamics. When mapped to the data, these features lead to a singular stationary equilibrium and no aggregate poverty trap (See, e.g., Buera et al., 2011; Buera and Shin, 2013, 2011; Buera et al., 2012, 2014).8

The quantitative models do exhibit the micro level poverty traps, however. Recall, that these models have discrete occupational choice decisions and larger fixed, entry, or technology upgrading costs that can give rise to a minimum profitable scale of production and minimum wealth thresholds for entrepreneurship. Individuals with identical circumstances and with similar wealth but on either side of the threshold tipping point can follow very different economic trajectories. One might wonder, why such thresholds cannot be overcome if self-financing is available. The basic intuition is that self-financing can be long and costly. People with wealth above the threshold will decide to accumulate and ultimately become entrepreneurs, while those below decide it is not worth it, and they run down their wealth over time. Hence, financial frictions can delay or completely eliminate entry for the productive poor.

3.2.3 Relevance for Aggregate Policy Analysis

Consider the relevance of these ideas for understanding the long-run impacts of large scale redistribution, a policy area in which these quantitative models have been applied. Buera et al. (2014) analyze the impacts of economy wide asset transfer programs to the poor. They

---

8In a recent paper, Buera et al. (2020b) endogenize productivity by modeling firms’ technology adoption decisions in a world with intermediates. Firms do not face financial frictions per se but idiosyncratic gross output distortions. Since adoption of others impacts the price of an aggregate intermediate, they find that an aggregate poverty trap exists in India, where adoption costs are high, but not in the U.S. However, India is in the good equilibrium rather than the poverty trap.
calibrate critical idiosyncratic shock processes by looking at firm size distributions, the levels and churning into and out of entrepreneurship, and volatility in income observed in South Asia. Although micro level poverty traps exist, no such poverty trap exists on the macro level, since shocks eventually move people through the distribution. This logic is used to predict that the aggregate short-run impacts of large-scale asset transfer programs would eventually dissipate. Nevertheless, the distribution exhibits substantial churning, so productive individuals lying near thresholds may have there economic trajectories dramatically altered. In contrast to the poor, productive entrepreneurs, necessity entrepreneurs are much more likely to simply return to their original state. In contrast, Buera et al. (2012) showed that a different policy, microcredit, could permanently impact the distribution, because it is a persistent policy.9 Moreover, at scale, microfinance can indirectly impact the poor through both wages and interest rates (Buera et al., 2012; Kaboski and Townsend, 2012; Breza and Kinnan, 2018)

4 Empirical and Data Needs for Future Research

The previous section touched on several advances in understanding and quantifying the impacts of financial frictions on firm growth, and ultimately macroeconomic growth. In this section, I discuss the data, data needs, and data limitations that direct current research on the subject.

4.1 Ideal Data

The existing theory tells us that we lack the ideal data to perfectly measure and discipline mechanisms even in the current models we have. For a static estimation, such ideal data would consist of the joint distribution of ability and wealth. For any type of policy analysis, however, one would need to think about how this distribution would endogenously and dynamically respond, which would require panel data to better discipline the dynamics of income and wealth accumulation, and potentially productivity. We would also know bilateral

---

9This result echoes that of Aghion and Bolton (1997), who showed that persistent redistribution was needed given financial frictions.
financial flows at the transaction level, terms and any other financing constraints. Finally, it would cover the full distribution including the wealthiest and the most productive as well as the wealth and potential productivity of non-entrepreneurs.

Clearly, such data do not exist, nor could they in many cases. Consider instead the available data in an advanced economy like the United States. For firms, the economic census surveys all (formal) firms every five years. Annual surveys cover all large firms and a representative sample of smaller firms. None of these are publicly available nor are they panel data. For panel data, the Compustat dataset is available, but it covers only public firms. For wealth and finances, the Survey of Consumer Finances tracks a small (less than 7000) representative sample of households every three years with detailed data on personal finances. In any case, none of these survey contain the wealthiest households, despite oversampling. For information on the absolute wealthiest tail of the distribution, researchers rely on imputing wealth from tax records (e.g., Saez and Zucman, 2016). These data can be used to yield current marginal distributions of firm productivity and distributions of wealth, but no data can inform the joint distribution or even correlations between the two. Moreover, the productivity data cannot inform us of the distribution of potential productivities of firms that do not exist.

The literature has addressed these challenges in multiple ways. Absent access to productivity data, firm size data have been used to discipline the productivity distribution. To address the lack of data on the productivities of those who don’t run firms, researchers have written models of selection. To approximate the wealth distributions, researchers have relied on simulations of models of dynamic savings and stochastic income disciplined by firm entry, firm growth, and measured moments of income distributions. Income microdata is more readily available than wealth data, but it still suffers from poor coverage in the far right tails. Nevertheless, simplifying distributional assumptions like Pareto or log normal productivity distributions, together with a moment about the concentration in the tails of the firm size and income distributions to calibrate the parameters, can help one reasonably approximate and discipline these distributions. Again, the wealthy and large firms play an especially important role in understanding aggregates.

A final area is the financial system and financial contracting itself. In the United States,
a host of financial data are available including publicly available data on stock and bond markets, call reports for regulated commercial banks, and the Federal Reserve System’s Flow of Funds data.\textsuperscript{10} At the microlevel, in principle, credit rating agencies aggregate debts, terms and records, although these data are not publicly available.

### 4.2 Data in Developing Countries

To truly and effectively assess the quantitative impacts of policies in developing countries, one would like to have data that reflect the countries themselves. Here things become more difficult as macro representative data are much more limited in less developed countries.

#### 4.2.1 Firm Data

Firm panel data are available for some of the larger economies: the Annual Survey of Chinese Industrial Enterprises (CIE) panel, the Annual Survey of Industries (ASI) from India, and the Relação Anual de Informações Sociais (RAIS), which is a matched employer-employee dataset of all formal firms in Brazil. Brazil also has their Annual Survey of Industry-Enterprises (PIA-Empresa) and surveys for services (Annual Survey of Services, PAS) and trade (Annual Survey of Trade, PAC), but the coverage is only panelized for the larger firms.

These datasets have their own limitations. Outside of manufacturing, coverage is weaker. In fact, the CIE covers only industrial firms. RAIS has complete coverage of formal firms but has not data on capital, value-added, or material usage data. The other datasets have limited coverage of smaller firms, and none of them (including RAIS) have coverage of informal firms. In some cases, one can supplement these to be representative of the economies overall but the panel aspect is then lost. Often the supplements are less frequent and periodic, however. For example, the Pesquisa de Economia Informal Urbana (ECINF) survey covers smaller and informal firms, but it was only conducted in 1997 and 2003. In Mexico, the National Survey of Microenterprises (Encuesta Nacional de Micronegocios or ENAMIN) is conducted roughly every two years, but the Economic Census, which covers only fixed establishments, is only

\textsuperscript{10}Flow of Funds are a set of aggregate financial accounts that are designed to mesh consistently with the real accounts in the National Income and Product Accounts. The accounts gives measures of asset and liability holdings of major sectors of the economy broken down by major asset types, durations, etc. The Flow of Funds data can be used to get aggregate savings, as well as sources and uses of financial capital.
conducted every five years. For publicly traded firms, Compustat Global covers roughly 90
countries, including publicly traded firms in many developing countries.

For the smaller, and often poorest, countries, such large comprehensive datasets are completely lacking. The research discussed above that used macro-representative micro data is overwhelmingly based on countries with higher incomes than those in sub-Saharan Africa, for example. Moreover, in the micro research studies on narrower samples, the results vary considerably by country and study. Clearly, much more research is needed, and that will require adapting imperfect data and newly arriving data.

Such data do exist, but the samples and panel structure are more limited, especially prior to 2000. Since 2000, the World Bank’s Enterprise Analysis Unit has collected the Enterprise Surveys, which are stratified, random samples that are designed to be nationally representative and stratified by enterprise size, sector, and geographic regions. Sample sizes are typically around 500 firms. They collect what one needs to estimate production functions, in principle, and they even contain some information on finance. The coverage of countries is quite broad, but the time span and number of rounds is small. For example, the typical African country has two surveys, 4-10 years apart. Moreover, these surveys omit the microenterprises and informal firms, both of which are of disproportionate importance to the poorest economies. For some countries, the standard Enterprise Survey is supplemented by surveys of microenterprises or informal enterprises, but these are even less frequent.

4.2.2 Income Data

Representative data on the distribution of income, and anything that can inform the distribution of wealth, is of high importance. On the firm side, the challenge is often to get the smallest firms. On the household side, the challenge is to get the wealthiest households. Indeed, wealth data themselves are difficult to come by, and so we must rely on the income and consumption data available in demographic censuses and household surveys. Census

---

11 The Center for the Study of African Economies, at STEG partner institution, Oxford University, has Manufacturing Enterprise Surveys for Tanzania and Ghana dating back to the 1990s, and for Kenya starting in the 2000s. See https://www.csae.ox.ac.uk/data.

12 A full list of countries, years, and samples are available here: https://www.enterprisesurveys.org/content/dam/enterprisesurveys/documents/methodology/DataDetails.xls. An explanation of data covered is available here: https://www.enterprisesurveys.org/content/dam/enterprisesurveys/documents/Indicator-Descriptions.pdf
data are available for most countries through IPUMS-International, although it is infrequent for all countries and less periodic for the poorer countries.

For the lower-to-middle income developing countries, nationally representative labor force surveys are common. Donovan et al. (2020) have assembled a large panel of countries. The NSS in India contains good data on both income and consumption jointly. In Brazil, the National Household Sample Survey (PNAD) is another annual household survey that yields labor and income data, but for consumption data, one must move to the Brazilian Consumer Expenditure Survey (POF). For most countries, the labor force surveys contain only income data, and typically not the right tail with the highest income.

For lower income countries, regular and representative labor force surveys are less common. However, the World Bank’s Living Standard Measurement Study (LSMS), in conjunction with national statistical offices, has produced a series of high-quality, nationally-representative household panel surveys. These surveys contain data on both consumption and income. In 2009, the program expanded into the LSMS-ISA (Integrated Surveys on Agriculture), which remained nationally representative but with a strong focus on agriculture. Landholdings and livestock are important indicator of wealth, especially rural wealth. Again, these data do not cover the wealthy tail, the tail which owns the largest domestically-owned establishments.

### 4.2.3 Financial Data

Data on the actual state of the financial system and access to financing are also difficult to come by in developing countries. At the level of measures of the depth and structure of financial aggregates, the World Bank’s Financial Structures Database, constructed by Beck et al. (2000), is a nice source. These data are assembled for a larger number of countries over various years using the International Monetary Funds’s International Financial Statistics, BankScope, and various country-specific sources.  

At a less aggregated level, Flow of Funds accounts are not typically available for developing countries. In many ways, the standard accounting apparatus, which aggregate the

---

13The World Bank’s Global Findex gives cross country data, based on underlying microsurvey data of the fraction of the population (by gender and grouped age) with access to various financial instruments.
household sector, the corporate sector, the government, etc. at a national level is much more appropriate for a well-integrated financial system, where costs and returns to capital are roughly equalized at the smaller scale. However, in a small country, the corporate sector may be getting sizable credit in aggregate, while certain types of businesses, certain industries, or certain regions are still quite pinched. Nevertheless, although the informal flows are considerable, typically central banks have the underlying data needed to construct formal flow and even at a more disaggregate basis.

In some ways, developing countries have better financial data than advanced economies. Comprehensive credit registries exist in many developing countries. Bustos et al. (2020) linked data from Brazil to spatial data on agricultural income and bank branches to study the impact of capital accumulation of industrialization. The aforementioned Cavalcanti et al. (2019) merges the data with firm panel data to connect credit spreads with firm growth. This is an avenue for more research.

In contrast, the data on financial inclusion are quite weak in developing countries, although effort and resources have been focused on the topic. What is really wanted is a measure of access to finance, as opposed to headcounts and fractions of people who hold formal savings accounts or loans as in the FinScope survey. Many people who have savings accounts do not use them to any measurable degree. Nor does the lack of a loan equate to a lack of access, e.g. Dupas and Robinson (2013). One chief reason for promoting financial inclusion, especially among poorer and remote borrowers is because they allow for easier risk sharing. However, research by Townsend (2010a) shows that much of the risk-sharing benefits are already gained if a relative has formal savings and/or loans. Interest rate spreads and fees may be a more useful measure of financial access, and in the past, bank proximity has been a useful measure.

One shining example of micro data is the work of Townsend and coauthors in Thailand. The backbone of this research is the Townsend Thai data, which has panel data from various regions in Thailand, and intensively covers household income, consumption, business activ-

---

14The aggregate nature of the Flow of Funds accounts in the U.S. was, by my own estimation, even an issue during the 2008 financial crisis, when in aggregate an early report claimed the idea that firms were facing a financing crunch was a “myth” since credit had not contracted on aggregate as reported by Chari et al. (2008).
ities, assets, savings, borrowing and lending. It is an extended panel with both annual and monthly survey components. These data are not macro representative – they suffer from the issues of missing the large firms and wealthy people in Bangkok, for example, but they have been used to construct “macro” accounts of village economies, including income and product accounts, current and capital accounts. Moreover, they have been combined with other existing data sources, spatial data on bank locations and road locations, nationally representative Socioeconomic Survey (SES), provincial and national aggregates, and even community level data for villages from the Community Development Office. Such creativity is needed in macro development.

5 Needed: Development Policy Research

In the above review of research on firms, financial frictions, and financial development, I touched on a number of different concrete policies, like credit bureaus, microfinance, and redistributive asset grant programs. Nevertheless, the research has only touched the tip of the iceberg in thinking about development policy.

5.1 Financial development policies

An empirical literature and quantitative literature both indicate the financial development is good for the economy but relatively little work has looked in concrete terms at how to promote financial development. This is a clear area for more research. In principle the models have the workhorse apparatus to consider a wide range of potential policy questions that developing countries face.

A starting point is getting a better sense of how important savings, credit, and insurance are to economic outcomes. Given the current models and apparatus, and the debates about self-financing and credit as substitutes, this is relatively low-hanging fruit.

Bank regulation is another important area for considering policy that promotes financial development. Of course, there are the standard policies of reserve and capital requirements, but there are also competition policies. The optimal size of banks is another important policy question. There are returns to scale and diversification for banks but there are also issues
of moral hazard when banks become “too big to fail”. In many countries, bank branching is regulated both to regulate competition and to promote access for under-served areas and population.

Whether or not to open to foreign banks and foreign capital flows is another essential macro regulatory policy that developing countries must face in trying to promote financial development. Foreign countries can be a flush source of capital, but this capital can also be fickle and can lead to financial fragilities, and crises in the case of outflows. Bankruptcy law is another area where the interplay of banks and borrowers becomes important. In a world with both informational frictions and exogenous shocks, bankruptcy policy can be important to ensuring both sufficient supply and demand for credit, but also stability in the system (Chatterjee et al., 2007). For all of these bank regulation questions, it would be necessary to introduce banks as agents into the existing models. Dean Corbae and d’Erasmo have a series of papers modeling banks with market power, (e.g., Corbae and D’Erasmo, 2018, 2019a,b) have developed a model of banks in competition but with market power to study regulations in the U.S. Integrating this structure into growth models with financial constrained entrepreneurs that might be promising for thinking about these issues for developing countries.

Tax policy is another way to potentially promote financial development. Recent work has emphasized the importance of a paper trail for formalizing relationships, and the VAT as a tax that gives incentives for formalization (Pomeranz, 2015).\textsuperscript{15} New models of input-output structure and firm-to-firm relationships enable easier analysis of such models (e.g., Baqae and Farhi, forthcoming; Boehm and Oberfield, forthcoming).

Corporate control and governance are also major policy issue. In a world of moral hazard of entrepreneurs, monitoring and costly state verification or a desire of entrepreneurs to retain control when nonpecuniary aspects are valued can lead to debt as the optimal financing mechanism (Townsend, 1979; Aghion and Bolton, 1992). Still, equity offers investors more corporate control. With proper corporate governance, equity markets can prevail; however, when boards do not have proper independence and oversight, both equity and lending markets can suffer. Family-owned and controlled businesses are common in developing countries,

\textsuperscript{15}Still, taxes themselves can distort supplier relationships (Gadenne et al., 2019).
and this may be partially the result of less developed financial markets. Family connections can enable internal capital markets that substitute for financial markets (Karaivanov et al., 2018). However, dynastic firm management can also lead to underperformance (Bertrand et al., 2008) and crony financial relationships (La Porta et al., 2003).

5.2 Alternatives to Financial Development

There is a literature that attributes financial underdevelopment to more underlying forces such as legal origins and property rights, which are difficult to affect through simple policies (e.g., La Porta et al., 1998). In such cases, second best policies must be considered. If financial frictions are bound to prevail, what other policies can mitigate their negative effects? Second-best policies include policies that cut against instincts in many ways, such as labor suppression (Itskhoki and Moll, 2019), anti-competitive practices, and deviations from free trade, all of which can enable productive firms to more readily accumulate profits and escape financial frictions.

To what extent, can the state compensate for poorly functioning financial markets through direct subsidized lending? Many countries have credit subsidy programs targeted toward particular industries and/or small and medium enterprises. The role of state and development banks, in particular, is another question for policy, due to the long history of poorly run state banks. There are, however, some examples of well-functioning state banks, such as Thailand Bank of Agriculture and Agricultural Cooperatives.

Another substitute for a poor financial systems is to attract foreign direct investment. In the context of the above models, multinationals may be best thought of as deep-pocketed, high productivity producers, whose capital income is expatriated back to the source country. They have the ability of producing jobs, helping bring and coordinate infrastructure investment, and highly efficient production. Much of the debate about foreign multinationals is how they impact the domestic firms, through spillovers, collaboration, and competition, and how they influence the political economy of a country. Assessing impacts of foreign multinationals and relevant multinational policies in the context of a world of financial frictions would be insightful. A caution is that informal alternatives to financial development typically already exist (Townsend, 1994). Formal finance can interact with informal networks, either
undermining the ability to sustain risk sharing arrangements (Kocherlakota, 1996; Ligon et al., 2002) or enabling further cooperation (Dupas et al., 2019).

5.3 Studying Financial Inclusion

I touched on the poor data and changing nature of financial inclusion, which remains a major policy goal. Yet, the costs and benefits of financial inclusion remain understudied, especially newer services. In addition to microcredit, other policies include micro savings accounts and even insurance schemes. Innovations in payment systems, like mobile money, is another area of financial inclusion. Yet the macroeconomic and distributional consequences of such innovations have been understudied.

In cases where financial inclusion has the potential to be beneficial, another important but nontrivial question is how to expand inclusion. For example, experimental evaluations of microcredit have often showed low take up (Banerjee et al., 2015). Low takeup may reflect that many people have no need for finance, but it may also indicate that certain forms of financial services better meet the needs of particular populations. Furthermore, in many places it appears that the lack of formal financial services is not simply a question of access but also of demand. In addition to standard transaction costs, issues of trust, financial literacy, and risk may limit the demand for financial services. Neither the empirical/experimental nor quantitative literature have examined the macro dimensions of these issues.

Understanding the costs and benefits of financial services to women is of particular importance for multiple reasons. First, women may have reduced access due to their societal status or other cultural factors. Second, in many LICs, the labor market is either less accessible to women or less desirable given household responsibilities. In these cases, financing other productive activities for women may be important. In principle, access to finance income generating activities can change bargaining power of women. While some of the benefits of inclusion in an equilibrium sense may be overstated, since many people effectively have access to financial services through friends and family, financial inclusion may impact both the financial system itself, the aggregate level and distribution of income and the stability of consumption needs to further considered.

In sum, there are many areas for future research in developing stronger guidance for
governments in how to develop their financial system, how to deal with weak financial systems, and whom should be included in the financial systems in order to encourage inclusive macroeconomic development.

6 Conclusion

This paper has reviewed the recent literature on how financial development hampers economic development through the impact of financial frictions on firms and entrepreneurship. There is exciting work that is complementary and coming from many approaches: theoretical, empirical, and quantitative, and there is much to gain from integrating approaches. However, research needs to be applied to developing countries to inform policy. Here the data limitations require flexibility, care, and creativity in extending this work to the developing countries, where it is most relevant. The ground is fertile, however, and many exciting avenues for future research exist to provide policy guidance for financial development itself, second best policies in the face of financial frictions, and promoting an inclusive financial system that benefits all.
References


