Abstract

The essays in this thesis study the implications of weak financial institutions for economic growth, allocation of resources, and economic development. Methodologically, the essays draw on a broad range of theoretical and empirical tools from both macro and microeconomics.

Many currently and previously developing countries have adopted industrial policies that push resources towards certain "strategic" sectors, and the economic reasoning behind such polices is not well understood. In Chapter 1, I construct a model of a production network where firms purchase intermediate goods from each other in the presence of credit constraints. These credit constraints distort input choices, thereby reducing equilibrium demand for upstream goods and creating a wedge between the potential sales ('influence') and actual sales by upstream sectors. I analyze policy interventions and show that, under weak functional form restrictions, the ratio between a sector's influence and sales is a sufficient statistic that guides the choice of production and credit subsidies. Using firm-level production data from China, I estimate my sufficient statistic for each sector and show that it correlates with proxy measures of government interventions into the sector. Using a panel of cross-country input-output tables and sectoral production tax rates, I show that the tax rates for developing countries in Asia also correlate with the model-implied intervention measure.

In joint work with Benjamin Roth, Chapter 2 offers a new explanation for why microcredit and other forms of informal finance have so far failed to catalyze business growth among small scale entrepreneurs in the developing world, despite their high return to capital. We present a theory of informal lending that highlights two features of informal credit markets that cause them to operate inefficiently. First, borrowers and lenders bargain not only over division of surplus but also over contractual flexibility (the ease with which the borrower can invest to grow her business). Second, when the borrower’s business becomes sufficiently large she exits the informal lending relationship and enters the formal sector—an undesirable event for her informal lender. We show that in Stationary Markov Perfect Equilibrium these two features lead to a poverty trap and study its properties. The theory facilitates reinterpretation of a number of empirical facts about microcredit: business growth resulting from microfinance is low on average but high for businesses that are already relatively large, and microlenders have experienced low demand for credit. The theory features nuanced comparative statics which provide a testable prediction and for which we establish novel empirical support. Using the Townsend Thai data and plausibly exogenous variation to the level of competition Thai money lenders face, we show that as predicted by our theory, money lenders in high competition environments impose fewer contractual restrictions on their borrowers. We discuss robustness and policy implications.
Motivated by the explosive growth of microfinance in India and the eventual collapse of the industry following the default crisis in 2010, my joint work with Daniel Green in Chapter 3 provides a theory that explains how institutional weakness in credit markets can fail to stimulate development even when there is ample credit supply. We show that when borrowers lack credible mechanisms to commit not to borrow further from other lenders in the future, not only does the increasing availability of lenders raise the interest rate on loans and reduce the amount of funds that entrepreneurs can borrow, but perversely it is those entrepreneurs with more profitable investment opportunities that will end up raising fewer investments precisely because they have stronger desires to seek out additional lenders in the future. This effect further discourages entrepreneurs from initiating the most efficient and productive endeavors, generating persistent underdevelopment.

Thesis Supervisor: Daron Acemoglu
Title: Elizabeth and James Killian Professor of Economics

Thesis Supervisor: Abhijit Banerjee
Title: Ford Foundation International Professor of Economics
Acknowledgments

I am extremely fortunate to have interacted with and learned from so many amazing individuals, without whom this thesis would not have been possible. First and foremost, I am indebted to my advisors Daron Acemoglu and Abhijit Banerjee. Daron has been a role model for my professional development—his knowledge, creativity, enthusiasm, and professionalism have been constant sources of inspiration ever since the day I took the courage to email him about taking me under his wings. Abhijit has been a wonderful mentor: his endless insights have had an enormous influence over my research, and it was his most generous support and his confidence in me that got me through the toughest times in graduate school. I am continuously amazed by the two of them, whom I hope to keep learning from in the future.

I am grateful to my advisor Rob Townsend, whose advice has shaped the second and third chapter of my thesis. I have benefited greatly from our interactions, for his infectiously cheerful personality and his generosity with his time. I look up to him for his love of economics and dedication to research.

I am also indebted to my former mentors, Matthew Gentzkow and Jesse Shapiro, who took me as a research assistant, introduced me to economics research, and provided continual support during graduate school. They hold every aspect of their work to the highest standard I can imagine, and I aspire to do the same for my own research. I especially thank Jesse for his insightful feedback on this thesis.

MIT has provided me with the ideal environment as a student of economics, and I have learned from many other faculty members here. Alp Simsek’s effective teaching attracted me to macroeconomics, and his sharp comments as well as the clarity of his style of research will continue to influence me—I regret not having spent more time bothering him. I would also like to thank Marios Angeletos, David Atkin, Arnaud Costinot, Esther Duflo, and Tavneet Suri for their encouragement, guidance and support during my graduate career.

I wouldn’t have survived graduate school without my wonderful peers. Ludwig Straub was the best gym buddy I could have hoped for. It took us only one gym session to realize that neither of us is interested in working out, and it is in these sessions that I’ve had some of the most fruitful economics discussions in graduate school. Daniel Green, who coauthored the third chapter of this thesis, has always been the first person I go to for feedback on my preliminary ideas. I have spent many memorable evenings playing board games with Vivek Bhattacharya, Greg Howard, Arianna Ornaghi, Bryan Perry, and Ben Roth (who coauthored the second chapter), discussing economics while trying to see through each other’s bluff. I look forward to the next time we gather. I would also like to thank Jie Bai, May Bunsupha, John Firth, Alex He, Yan Ji, Gabriel Kreindler, Yuhei Miyauchi, Pooya Molavi, Scott Nelson, Harry Pei, Jeff Picel, Yu Shi, Gea Hyun Shin, and Linh Tô, for our discussion of economics and for friendships that made my overall graduate school experience
much more enjoyable. I hope we will continue to be friends for a long time.

Finally I would like to thank my family. Meeting my wife, Yan Leng, was the best thing that happened to me at MIT. I dedicate this thesis to her and my parents, for the sacrifices they made for me.
List of Figures

1-1 Illustration of the Vertical Production Economy ........................................ 25
1-2 Density of sectoral private return $(\phi^K)^{-1}$ and social return $SR_i^K$ ....... 55
1-3 Industries with higher sales gaps have more SOEs ...................................... 62
1-4 Average Sectoral Log(Sales Gap) For LDCs ..................................................... 69

2-1 Patterns of Default by Wealth Level ............................................................. 197
2-2 Probability Household Borrows from Money Lender by Wealth Level .............. 197
2-3 Probability Money Lender Uses Restrictive Collateral by Wealth Level ............ 198
2-4 Coefficients on $invesize_v \times yeart$ .......................................................... 198

3-1 Timing of the Model ......................................................................................... 209
3-2 The Lending Market Game at $s = 1$ ............................................................. 209
3-3 Static Representation of Lending Market Equilibrium. $\phi > 1$ indicates an increase in marginal returns to new investment. This shifts the borrower's loan aggressiveness best response curve upward but does not change the lender best response curve, which is only a function of $q$ .......................................................... 217
3-4 Expected Investment in the Dynamic Lending Game ........................................ 221
3-5 Expected Investment and Expected Debt ....................................................... 222
3-6 Endogenous Project Choice ............................................................................ 224
List of Tables

1.1 Summary Statistics ................................................................. 60
1.2 Top and Bottom 15 Industries Ranked by Sales Gap .................... 62
1.3 SOEs have higher Debt-To-Capital ratios and pay lower interest rates 65
1.4 Private firms in sectors with high sales gaps receive more external loans 66
1.5 Private firms in sectors with higher sales gaps pay lower interest rates 67
1.6 SOEs have higher value-added shares in industries with high sales gaps but not in those that are constrained or have high private returns to spending on capital. 70
1.7 Top and Bottom 5 Manufacturing Industries Ranked by Sales Gap in China 74
1.8 On average, LDCs have higher subsidies in sectors with larger sales gap 80
1.9 On average, Asian LDCs have higher subsidies in sectors with larger sales gap 82
1.10 Countries in the OECD IO Table Dataset .................................. 85
1.11 Non-Public Sectors in the OECD IO Table Dataset ..................... 86
1.12 Regression of sectoral subsidies on sales gap, by country .......... 114
1.11.1 Replication of main results in section 1.3 using $\xi^{NX}$ as the sales gap measure 115
1.11.2 Replication of main results in section 1.3 using $\xi^{M}$ as the sales gap measure 116
1.11.3 Replication of Tables 1.3 to include the sample of firms that report zero interest payments .......................................................... 117
1.11.4 Replication of Tables 1.5 to include the sample of firms that report zero interest payments .......................................................... 118
1.11.5 Replication of Table 1.6 using SOE share of industry gross value-added and industry revenue ......................................................... 119
1.11.6 Replication of main results in section 1.3, dropping joint-ventures and firms with foreign ownership ............................................. 120
1.11.7 Number of firms in each industry by firm type ........................ 120
1.11.8 Specification test: hypothetical wedges on labor and material inputs 121
1.11.9 Replication of main results in section 1.3, assuming firm-production functions have constant output elasticity for capital inputs ............................................. 122
2.8.1 Default is U-Shaped in Income in Field et. al. Data ............... 204
2.8.2 Village Covariates Correlated With Inverse Size ................. 205
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.8.3</td>
<td>Restrictive Loans are Correlated with Lower Income Expectation</td>
<td>206</td>
</tr>
<tr>
<td>2.8.4</td>
<td>Restrictive Loans Are Correlated with Lower Interest Rate</td>
<td>207</td>
</tr>
<tr>
<td>2.8.5</td>
<td>Main Comparative Static on Village Fund Intensity</td>
<td>208</td>
</tr>
<tr>
<td>2.8.6</td>
<td>Parallel Trends Prior to 2002</td>
<td>209</td>
</tr>
<tr>
<td>2.8.7</td>
<td>Whether a Household Borrows from Money Lenders Regressed on ( post \times invsize )</td>
<td>210</td>
</tr>
<tr>
<td>2.8.8</td>
<td>Total Household Borrowing from Money Lenders Regressed on ( post \times invsize )</td>
<td>211</td>
</tr>
<tr>
<td>2.8.9</td>
<td>Heckman Selection Model</td>
<td>212</td>
</tr>
<tr>
<td>2.8.10</td>
<td>Placebo Test: Main Regression with Population of Borrowers From Neighbors Rather Than Money Lenders</td>
<td>213</td>
</tr>
<tr>
<td>2.8.11</td>
<td>Placebo Test: Main Regression with Unrestrictive Collateral as Outcome Variable</td>
<td>214</td>
</tr>
<tr>
<td>2.8.12</td>
<td>Main Regression with Interest Rates as Outcome Variable</td>
<td>215</td>
</tr>
<tr>
<td>2.8.13</td>
<td>Robustness Check: Main Comparative Static Without Restriction on Village Size</td>
<td>216</td>
</tr>
<tr>
<td>2.8.14</td>
<td>Robustness Check: Main Comparative Static Using Data Only From 1999-2004</td>
<td>217</td>
</tr>
<tr>
<td>2.8.15</td>
<td>Robustness Check: Main Comparative Static with Data Driven Definition of Restrictiveness</td>
<td>218</td>
</tr>
</tbody>
</table>
Chapter 1

Industrial Policies and Economic Development

1.1 Introduction

Industrial policies are broadly defined as the selective interventions that attempt to alter the structure of production towards certain sectors. Such policies are not only widely adopted in developing countries today, but also played a prominent role in the developmental stage for many now-advanced economies. Prime historical examples of industrial policies include Japan in the 1950s and 1960s and South Korea and Taiwan in the 1960s and 1970s. In all of these cases, the government heavily promoted "strategic" upstream sectors that supply to many others sectors. A wealth of policy instruments was adopted during these periods, including various forms of tax incentives and subsidized credit, and in the case of Taiwan, direct state involvement in production. In Korea, the explicit industrial movement was termed the "Heavy-Chemical Industry" drive, and for almost a decade firms in selected industries received policy loans with significantly reduced interest rates (Amsden 1989, Woo-Cumings 2001). Total policy loans directed towards the targeted sectors accounted for 45% of the total domestic credit of the banking system in 1977 (Hernandez 2004). Many of the largest manufacturing conglomerates in Korea today originated during this era.

By their nature, these policies seek to affect the development of the aggregate economy through selective intervention in a few sectors. Understanding the effects of such intervention therefore requires modeling the linkages among sectors in the economy. Moreover, the frequent use of subsidized or targeted loans suggests financing constraints play an important role in the design of these policies. Motivated by these facts, this paper develops a framework for studying optimal industrial policy in a general equilibrium setting with financial frictions and network linkages among sectors.

In my model, production requires factor inputs as well as intermediate goods produced by other
sectors, and firms face credit constraints when purchasing some of these inputs for production. These credit constraints distort input choices and endogenously affect sectoral input-output linkages, thereby reducing demand for upstream goods that are subject to constraints. In equilibrium, the constraints generate a wedge between the total sales of the affected upstream sectors and the elasticity of aggregate output with respect to sectoral Hicks-neutral productivity shocks. This elasticity, known as the sectoral “influence” in the production networks literature, can be interpreted as the potential sectoral sales absent market imperfections (Hulten 1978). I analyze policy interventions and show that, under weak functional form restrictions, the ratio between a sector’s influence and sales—which I refer to as the sectoral “sales gap”—is a sufficient statistic that summarizes the inefficiencies in the input-output network and could guide policy interventions that expand sectoral production. Specifically, I show that starting from a decentralized equilibrium without distortionary taxes, a sector’s sales gap captures the ratio between social and private marginal return to spending resources in the sector on production inputs and on credit. Moreover, if production functions are iso-elastic, the same sufficient statistic captures the optimal sectoral subsidies to labor, which is the value-added input in the model. These results are potentially surprising because sectors with the highest sales gaps are not necessarily the sectors in which firms are most constrained; instead, they are upstream sectors that directly or indirectly supply to many constrained downstream sectors. In fact, my results imply that even if the private returns to credit are equalized across all firms in the economy, a benevolent planner might still want to direct credit to upstream sectors in order to improve production efficiency.

The sales gap is a sufficient statistic for network inefficiencies because while sales capture the relative sectoral size under equilibrium production in the presence of frictions, influence captures the relative sectoral size under optimal production. The distance between the two vectors thus reveals a direction in which production efficiency can be improved. This finding can be viewed as an “anti-network” result similar to Hulten’s: as long as we know the sales gap—the difference between a sector’s potential and actual sales—knowledge of the underlying frictions in the input-output system becomes irrelevant for welfare analysis.

I conduct two distinct empirical exercises to estimate sales gap and examine its correlations with proxy measures of government interventions. The first exercise focuses on China, whose socialist roots and strong legacy of state intervention makes it a particularly interesting setting to apply my results. Relying on firm-level manufacturing census, I estimate firm production elasticities and the distribution of credit distortions for the manufacturing sectors, and I use these estimates to compute sectoral influence and sales gap based on the observed Chinese input-output table. I find that private firms in sectors with higher sales gaps tend to receive more external loans and pay lower interest rates, and that the sectoral presence of Chinese State-Owned Enterprises (SOEs) is heavily directed towards sectors with higher sales gaps. My theory suggests that these selective interventions can enhance welfare by effectively subsidizing upstream production and potentially ameliorating the network inefficiencies. My findings therefore allow for a positive reappraisal of the
selective state interventions in China and provide a counterpoint to the prevailing view (e.g. Song et al. 2011) that SOEs are a sign of sectoral inefficiency.

My second empirical exercise compares across countries. Using a panel of cross-country input-output tables, I construct the sales gap measure for a set of developing countries based on the input-output tables from a set of developed countries, adopting a strategy that is similar in spirit to Rajan and Zingales (1998) and Hsieh and Klenow (2009). I show that, as a group, developing countries tend to have higher sales gaps in tertiary and heavy manufacturing sectors and lower sales gaps in primary and light industrial sectors. Moreover, I show that the sectoral sales gaps of a set of developing countries in Asia strongly and positively correlate with a measure of sectoral subsidies adopted by these countries. The pattern is largely absent or even reversed in developing countries from the other continents, which on average have had worse economic performances in recent years than their Asian counterparts. These results are consistent with the hypothesis that governments in countries with strong economic performances are better at understanding the network distortions and are adopting policies to address them.

Literature Review  My paper is related to a large body of development macro literature on the misallocation of resources, including Banerjee and Duflo (2005), Jeong and Townsend (2005), Restuccia and Rogerson (2008), Hsieh and Klenow (2009), Banerjee and Moll (2010), Song et al. (2011), Buera et al. (2011), and Rotemberg (2014), among many others. The broad purpose of this literature is to study the implications of micro-level financial frictions on aggregate productivity. My paper draws on this literature but provides a different focus. Rather than studying how financial constraints distort the efficient use of resources within a sector, I study how constraints endogenously affect input-output linkages and distort the relative size of sectors, generating the misallocation of resources across sectors in a production network.

The importance of sectoral linkages for economic development was first pointed out by Hirschman (1958), who argues that industrial policies should target and promote sectors with the strongest linkages. His work has inspired an early and substantial development economics literature that aims to measure the Hirschmanian linkages and study their relationships with economic performance and industrial policies, including Chenery and Watanabe (1958), Rasmussen (1965), Yotopoulos and Nugent (1973), Chenery et al. (1986), Jones (1976), and Shultz (1982), among others. My paper revisits this topic using a model with neoclassical microfoundations to formalize the implications of linkages for industrial policies.

My modeling approach embeds cross-sector input-output linkages into a static version of the competitive entry model with the convex-concave technologies of Hopenhayn (1992) and Hopenhayn and Rogerson (1993). The model sits squarely within the class of generalized Leontief models as defined in Arrow and Hahn (1971, pp. 40, Leontief economy). This class of generalized Leontief models has been extensively studied in the early general equilibrium literature, including Hulten
(1978), who shows that without market imperfections and under aggregate constant returns to scale, sectoral influence is equal to sales, an equivalence that is broken in my model due to financial frictions.

A modern revival of this Leontief input-output approach, often referred to as the "production networks" literature, imposes functional forms on generalized Leontief models and more explicitly studies how productivity shocks transmit through input-output linkages. Key contributions to this literature include Long and Plosser (1983), Horvath (1998, 2000), Dupor (1999), Shea (2002), and Acemoglu et al. (2012). Several papers in this literature embed financial frictions into production networks: Jones (2013) and Bartelme and Gorodnichenko (2015) model financial frictions through implicit wedges or distortions in factor prices à la Hsieh and Klenow (2009), and Altinoglu (2015) and Bigio and La'O (2016) model frictions through working capital constraints. These papers aim to characterize how linkages amplify sectoral financial frictions and study their implication on aggregate output. Relatedly, Baqace (2016) studies a production network model with monopolistic markups, which also create wedges between marginal product and marginal cost of production inputs. Market imperfections in the models of these papers also break Hulten's equivalence theorem.

My theoretical analysis differs substantively from those offered by the current production networks literature. First, because the decoupling of influence and sales is at the heart of my analysis, I conduct a detailed characterization of how sectoral constraints and network structure affect the sales gap. Second, I show that under aggregate constant returns to scale, sectoral size is proportional to influence under optimal production, even though it is proportional to sales under equilibrium production in the presence of constraints. I further show that their ratio, the sales gap, exactly captures the ratio between social and private marginal returns to spending productive resources in a sector, thus providing a direction in which production can be improved through policy intervention. These results do not rely on the Cobb-Douglas or the constant elasticity-of-substitution assumptions imposed by the production networks literature.

Baqace (2015) and Acemoglu et al. (2016) observe that in a production network under Cobb-Douglas technology assumption, productivity shocks travel downstream through input-output linkages from suppliers to buyers, while demand shocks travel upstream. My paper shows that these results can be generalized without the specific functional form assumptions, and I apply these intuitions to show how credit constraints affect allocations and distort the relative size of sectors. First, as an application of the non-substitution theorem by Samuelson (1951), demand shocks have no effect on equilibrium prices in a generalized Leontief model and affect equilibrium quantities only through backward linkages or, in other words, by traveling upstream. Second, even without the Cobb-Douglas assumption, productivity shocks in a sector travel through forward linkages, affecting the unit cost of production hence equilibrium prices of downstream buyers. Equilibrium prices of upstream sectors are unaffected, and output quantities in upstream sectors change in response to downstream productivity shocks only through the changes in demand induced by these shocks. Lastly, financial frictions in a generalized Leontief model serve both as a productivity shock and
a shock to intermediate demand that emanates from the constrained sectors. The productivity shock aspect of financial frictions propagates downstream by lowering aggregate output, while the demand shock aspect propagates upstream and suppresses the relative size of upstream sectors.

My model prominently features pecuniary externalities and thus relates to the literature on the inefficiency of general equilibrium with prices in additional constraints, including seminal work by Greenwald and Stiglitz (1986) and Geanakoplos and Polemarchakis (1986). More broadly, my policy analysis leverages the fact that in the presence of input-output linkages, policy instruments need not directly target the source of distortions in order to improve welfare. My analysis therefore relates to the second-best literature initiated by Lipsey and Lancaster (1956) and more recently contributed by Farhi and Werning (2013), Kilenthong and Townsend (2016), and Korinek and Simsek (2016), among others.

While I provide closed-form solutions for second-best policies under Cobb-Douglas assumptions, my main results study welfare changes in response to a marginal change in production subsidies. This approach is related to a series of papers in public finance literature, including Ahmad and Stern (1984), Deaton (1987), and Ahmad and Stern (1991), that study the welfare effect of marginal tax reforms.

The empirical setting of my analysis builds on a broad empirical and policy literature on state intervention and industrial policies, including Pack and Westphal (1986), Chenery et al. (1986), Amsden (1989), Wade (1989), Wade (1990), Westphal (1990), Page (1994), Pack (2000), Noland (2004), and more recently, Rodrik (2004, 2008) among others. Compared to the Computable General Equilibrium approach adopted by some papers in the policy literature, such as Dervis et al. (1982) and Robinson (1989), my work is microfounded as I explicitly model firm-level incentives and their behavior under credit constraints.

My first empirical exercise focuses on China and relates to a large literature that studies the growth experience of the country, including Brandt et al. (2008), Song et al. (2011), Zhu (2012), Bai et al. (2014), Storesletten and Zilibotti (2014), Aghion et al. (2015), and Hsieh and Song (2015). Most notably, I borrow from Song et al. (2011) in modeling SOEs as unconstrained profit maximizers, an assumption that plays a central role in this empirical analysis. The observation that Chinese SOEs are more present in upstream sectors has also been made by Li et al. (2015), who adopt the “upstreamness” measure by Antras et al. (2012).

My second empirical exercise uses cross-country input-output tables to test the relationship between the sales gap measure and a measure of sectoral tax rates in developing countries. I use observed input-output tables for developed countries to predict undistorted production technologies for developing countries, an exercise that is similar in spirit to Rajan and Zingales (1998) and Hsieh and Klenow (2009). Bartelme and Gorodnichenko (2015) conduct a similar exercise to infer unconstrained input-output production technologies, though they examine a different empirical relationship, one that is between aggregate TFP and a measure of total linkages across industries.
The rest of the paper is organized as follows: Section 2 provides the theoretical results, Section 3 conducts the empirical exercise based on firm-level Chinese manufacturing data, Section 4 conducts the cross-country analysis, and section 5 concludes.

1.2 Theory

1.2.1 Model Setup

Economic Environment There is a representative consumer who consumes a unique final good (with price normalized to 1) with non-satiated preferences and supplies labor $L$ inelastically. There are $S$ intermediate production sectors in the economy, each producing a differentiated good that is used both for intermediate production and also the production of the unique final good. I will refer to the output of sector $i$ as good $i$ and refer to the $S$ goods altogether as “intermediate goods”.

The final good is produced competitively by combining intermediate goods under production function

$$F(Y_1, \ldots, Y_S)$$

where $Y_i$ is the units of good $i$ used for final production. I assume $F(\cdot)$ is differentiable, has constant-returns-to-scale, and is strictly increasing and jointly concave in its arguments.

Production of intermediate goods is modeled as a two-stage entry game. In the first stage, a large measure of identical, risk-neutral, and atomistic potential entrants freely choose whether to set up a firm in any sector, taking the expected profit and cost of entry as given. In the second stage, firms that have entered in sector $i$ produce an identical and perfectly substitutable good $i$. To build a firm in sector $i$, an entrant $\nu$ pays a fixed cost $r_i$ units of the final good and acquires a production technology

$$q_i(\nu) = h_i \cdot z_i(\nu) f_i(\ell_i(\nu), m_{i1}(\nu), \ldots, m_{iS}(\nu)),$$

where $\ell_i(\nu)$ is the amount of labor employed by firm $\nu$ in sector $i$, $m_{ij}(\nu)$ is the amount of good $j$ used as intermediate inputs for production of good $i$ by firm $\nu$, and $z_i(\nu)$ captures firm-specific Hicks-neutral productivity. Lastly, $h_i$ is a sector-wide Hicks-neutral productivity shock common to all firms in sector $i$, which is introduced for notational purposes and is normalized to $h_i \equiv 1$ unless explicitly noted.

The model formulation implicitly assumes no joint production—each industry produces only one good. We make the following assumptions on $f_i$ and $F$:

**Assumption 1.1.** Production functions $f_i$ and $F$ are continuously differentiable and strictly concave. Furthermore,
a) $F(\cdot)$ satisfies the Inada conditions:

$$\lim_{Y_i \to 0} \frac{\partial F(Y_1, \ldots, Y_i, \ldots, Y_S)}{\partial Y_i} = \infty, \quad \lim_{Y_i \to \infty} \frac{\partial F(Y_1, \ldots, Y_i, \ldots, Y_S)}{\partial Y_i} = 0.$$  

b) $f_i(0, m_{i1}, \ldots, m_{iS}) = 0$ and $\frac{\partial f_i(t_i, m_{i1}, \ldots, m_{iS})}{\partial t_i} > 0$ at all input levels. That is, every firm needs labor to produce and output is always strictly increasing in labor.

Financial Constraints Financial frictions in this network economy are modeled as pledgeability constraints faced by firms in the intermediate goods sectors. I assume the cost of a subset of production inputs has to be paid before production takes place, and each entrepreneur $\nu$ has an exogenous amount of expendable funds $W_i(\nu)$. Formally, for each firm $\nu$ in industry $i$, there is a subset of inputs $K_i \subset \{1, \ldots, S\}$ that is subject to constraints of the following form:

$$\sum_{j \in K_i} p_j m_{ij} \leq W_i(\nu) \tag{1.1}$$

where $p_j$ is the price of good $j$, and $m_{ij}$ is the amount of good $j$ used for production. I use $K_i$ to denote the set of constrained intermediate inputs and $X_i$ to denote the set of unconstrained inputs, with $K_i \cup X_i = \{1, \ldots, S\}$. In this static production model, the left-hand side of the financial constraint (1.1) can be interpreted as an upfront payment requirement on certain inputs, before the firms make sales and are able to recover the input expenditures. The right-hand side of the constraint captures the total available funds to cover such upfront costs, which can be interpreted as the sum of entrepreneurial wealth and the total bank credit available to the entrepreneur to purchase the constrained inputs.

Inputs in $K_i$ that are subject to the constraint and can be thought of as capital goods (e.g. machinery, equipments and computers) or services that can be subject to hold-up problems (such as outsourced R&D services), for which trade credit is difficult to obtain and costs must be incurred upfront. The unconstrained inputs in $X_i$ can be thought of as material or commodity inputs—such as intermediate materials for the production of consumer goods (e.g. textiles), commoditized services, and energy inputs—for which trade credit is more available (Fisman 2001) such that the input cost can be paid after production is carried out.

The fact that labor input is unconstrained is not important for my theoretical results: the same sufficient statistic will capture the ratio between social and private marginal return to spending additional resources on any production input, including labor, whether or not the input is constrained. On the other hand, when I apply my model to data in sections 3 and 4, I take the empirical stance that labor is unconstrained. This assumption is motivated by the empirical evidence that firms in developing countries do not seem to be constrained in labor choices (Cohen 2016, De Mel et al. 2016). Note also that that the fixed cost of entry $\kappa_i$ does not appear in constraint (1.1), but this is
without loss of generality: I can always relabel the amount of exogenous expendable funds as $\bar{W}_i$ and define $W_i \equiv \bar{W}_i - \kappa_i$. Similarly, we can also reinterpret the constraint (1.1) as requiring only a fraction of the cost of capital inputs to be paid upfront by relabeling $W_i$ with a multiplicative constant.

In Appendix 1.8 I consider several different formulations of financial frictions. Appendix 1.8.1 reformulates the model with financial frictions in the form of a monitoring cost that is linear in the amount of credit delivered. The linear monitoring cost creates an exogenous wedge between marginal product and marginal cost of inputs, similar to the implicit wedges in Hsieh and Klenow (2009), Jones (2013), and Bartelme and Gorodnichenko (2015), and our results survive in that environment. I relax the constraint formulation (1.1) in Appendices 1.8.2 and 1.8.3 by successively introducing input-specific requirement for upfront payment (with the left-hand-side of constraints taking the form of $\sum_{j \in S} \eta_{ij}(\nu) p_j m_{ij}$ for $\eta_{ij}(\nu) \in [0, 1]$) and partial pledgeability of revenue (by introducing $\delta_i(\nu) p_i q_i(\nu)$ to the right-hand-side of constraints). I show that all of my theoretical results survive when production inputs have varying degrees of upfront-payment requirement $\eta_{ij}(\nu)$. When revenue pledgeability is also introduced, the constraint formulation nests the pledgeability constraints in Bigio and LaO (2016) and my results still hold if within-sector firm heterogeneity is removed, an assumption maintained by other papers in this literature.

My theory focuses on intermediate production, and I assume the final good producer operates without any credit constraints.

**Firm’s Profit Maximization Problem** Firms choose inputs in order to maximize profit subject to the credit constraint in (1.1):

$$\max_{\{m_{ij}\}} \max_{j=1}^S p_i q_i(\nu, \ell, \{m_{ij}\}) - \sum_{j=1}^S \eta_{ij}(\nu) p_j m_{ij} - w\ell \quad \text{subject to (1)}.$$ 

**Free Entry** Before production takes place, there is a large (unbounded) pool of prospective entrants into any industry, and all potential entrants are identical ex-ante. After incurring the fixed cost of entry $\kappa_i$ units of the final good, firms independently draw Hicks-neutral productivities $z_i(\nu)$ and expendable funds $W_i(\nu)$ from a sector-specific distribution with a compact, non-negative support and cumulative distribution function $\Phi_i(\cdot)$. Because all same-sector firms with identical productivity and wedges make the same allocation choice, I abuse the notation and use $\nu$ as the index for both the random draws of $(z_i(\nu), W_i(\nu))$ and also for the firm with these draws.

To make entry decisions, prospective entrepreneurs form rational expectations on the variable profits $\pi_i(\nu) \geq 0$, the maximand of $(P_{\text{firm}})$. The expected profit net of fixed cost in sector $i$ is $(\mathbb{E}_\nu [\pi_i(\nu)] - \kappa_i)$. If this value were negative, no firm would want to enter. In any equilibrium
where entry is unrestricted, an assumption I maintain, this value cannot be strictly positive, hence

\[ \kappa_i = E_\nu [\pi_i (\nu)]. \quad (1.2) \]

**Equilibrium**

**Definition 1.1.** A decentralized equilibrium is a collection of prices \( \{p_i\}_{i=1}^S \), wage rate \( w \), measure of firms \( \{N_i\}_{i=1}^S \), firm-level allocations \( \{\ell_i (\nu), m_{i1} (\nu), \cdots, m_{iS}, q_i (\nu)\}_{i=1,\cdots,S} \), production inputs for the final good \( \{Y_i\}_{i=1}^S \), aggregate consumption \( C \), net aggregate output \( Y \), and aggregate labor supply \( L \) such that

a) The representative consumer maximizes utility subject to his budget constraint, such that:

\[ wL = C. \]

b) A firm \( v \) in each sector \( i \) solves the constrained profit maximization problem \( (P_{\text{firm}}) \), taking wage rate \( w \), prices \( \{p_i\}_{i=1}^S \), its own productivity \( z_i (\nu) \), and expendable funds \( W_i (\nu) \) as given.

c) Free-entry drives ex-ante profits to zero in all sectors such that equation (1.2) holds.

d) Production inputs for the final good solve the profit maximization problem of the final producer

\[ (Y_1, \cdots, Y_S) = \arg\max_{\{\bar{Y}_i\}} F (\bar{Y}_1, \cdots, \bar{Y}_S) - \sum p_i \bar{Y}_i. \quad (1.3) \]

e) All markets clear:

\[ (\text{labor}) \quad L = \sum_i L_i \quad (1.4) \]

\[ (\text{interm. good } j \text{ for all } j) \quad Q_j = Y_j + \sum_i M_{ij} \quad (1.5) \]

\[ (\text{final good}) \quad Y = F (Y_1, \cdots, Y_S) - \sum_i \kappa_i N_i, \quad (1.6) \]

where capital case letters \( L_i, M_{ij}, Q_j \) denote total sectoral quantities:

\[ L_i \equiv N_i \int_{\nu} \ell_i (\nu) \, d\Phi_i (\nu) \]

\[ M_{ij} \equiv N_i \int_{\nu} m_{ij} (\nu) \, d\Phi_i (\nu) \]
\[ Q_i = N_i \int_{\nu} q_i(\nu) d\Phi_i(\nu). \]

f) The net aggregate output equals consumption:

\[ Y = C. \]

Before I introduce government expenditure in section 1.2.4, net aggregate output \( Y \) always equals aggregate consumption \( C \), and I use the two terms interchangeably.

**Example \( \mathcal{E} \)** There is no closed-form solution for equilibrium allocations without additional functional form assumptions. To make the discussion concrete, I sometimes refer to a specific three-sector example that is nested under my model. I refer to the example as \( \mathcal{E} \) and the setup is as follows. There are \( S = 3 \) intermediate production sectors in the economy, and these sectors form a vertically connected production network: good 1 is produced upstream using labor only, good 2 is produced by combining good 1 and labor, and good 3 is produced downstream by combining good 2 and labor. I assume the final good is produced linearly from the downstream good 3:

\[ F = Y_3. \]

I remove heterogeneity across firms within a sector, and I drop the firm index \( \nu \) to simplify notation. The firm-level production functions take the iso-elastic form:

\[ q_1 = \ell_1^{\alpha_1}, \quad q_2 = \ell_2^{\alpha_2} m_{21}^{\sigma_2}, \quad q_3 = \ell_3^{\alpha_3} m_{32}^{\sigma_3}, \]

where \( q_i \) is the firm-level output, \( \ell_i \) is the labor input for production, and \( m_{i,i-1} \) is the amount of good \( i-1 \) used by a firm in the production of good \( i \). I normalize \( \alpha_1 \equiv \alpha_i + \sigma_i \) for sectors \( i = 2, 3 \) so that the concavity of production is constant across all three sectors (to avoid carrying additional constants and obfuscating notation for this example). I normalize firm-level productivity to \( z_i \equiv 1 \).

All intermediate goods are subject to credit constraints. I also assume \( W_i \) is constant for all firms in sector \( i \):

\[ p_1 m_{21} \leq W_2, \quad p_2 m_{32} \leq W_3. \quad (1.7) \]

The flow of inputs and outputs in the network is represented in figure 1-1.
1.2.2 Equilibrium Characterization

**Firm-level Allocations** Under Assumption 1.1, the solution to firm $\nu$'s profit maximization problem post-entry ($P_{\text{firm}}$) is characterized by the first-order conditions with respect to inputs, which can be re-arranged into the following set of expenditure share equations:

$$\frac{w_{i}(\nu)}{p_{i}q_{i}(\nu)} = \frac{\partial \ln f_{i}(\nu)}{\partial \ln \ell}$$  

(1.8)

$$\frac{p_{j}m_{ij}(\nu)}{p_{j}q_{i}(\nu)} = \frac{\partial \ln f_{i}(\nu)}{\partial \ln m_{ij}} \text{ for all } j \in X_{i}$$  

(1.9)

$$\frac{p_{j}m_{ij}(\nu)}{p_{j}q_{i}(\nu)} = \varphi_{i}^{K}(\nu) \frac{\partial \ln f_{i}(\nu)}{\partial \ln m_{ij}} \text{ for all } j \in K_{i}$$  

(1.10)

The first two equations are standard: because labor and intermediate goods $j \in X_{i}$ are unconstrained, their expenditure shares $\frac{w_{i}(\nu)}{p_{i}q_{i}(\nu)}$ and $\frac{p_{j}m_{ij}(\nu)}{p_{j}q_{i}(\nu)}$ are equal to the respective output elasticity of the firm production functions evaluated at equilibrium quantities of inputs. On the other hand, this equivalence breaks down for intermediate goods $j \in K_{i}$ that are subject to credit constraints, reflected by firm-specific wedge $\varphi_{i}^{K}(\nu) \leq 1$, which shows up in the expenditure share equation for the constrained inputs. When the credit constraint binds, $\varphi_{i}^{K}(\nu) < 1$, and it distorts input expenditure downwards relative to the efficient level.

For each firm, the wedge $\varphi_{i}^{K}(\nu)$ is pinned down by equilibrium prices and firm-specific random draws. The inverse wedge $\left(\varphi_{i}^{K}(\nu)\right)^{-1}$ can be interpreted as the firm's private return to spending on constrained inputs, while $\left(\varphi_{i}^{K}(\nu)\right)^{-1} - 1$ captures the marginal gains of having additional working capital and is the interest rate the firm is willing to pay to obtain credit.
Definition 1.2. The private return to spending on input $j$ for firm $\nu$ in sector $i$ is the ratio between the marginal product and marginal cost of input $j$:

$$PR_i^j(\nu) \equiv \frac{p_i \frac{\partial q_i(\nu)}{\partial m_{ij}}}{p_j}.$$

Similarly, the private return to spending on labor is

$$PR_i^l(\nu) \equiv \frac{p_i \frac{\partial \ell_i(\nu)}{\partial \ell_i}}{w_i}.$$

Lemma 1.1. Consider the inverse of the wedge on intermediate input in sector $i$, $(\varphi_i^K(\nu))^{-1}$.

a) Let $\eta_i(\nu) \geq 0$ be the Lagrange multiplier on the financial constraint (1.1) for the firm $\nu$'s profit maximization problem ($P_{firm}$). We have

$$(\varphi_i^K(\nu))^{-1} = 1 + \eta_i(\nu).$$

b) $(\varphi_i^K(\nu))^{-1}$ captures the private return to spending on constrained inputs:

$$(\varphi_i^K(\nu))^{-1} = PR_i^j(\nu) \text{ for all } j \in K_i.$$

c) $(\varphi_i^K(\nu))^{-1} - 1$ captures the firm's marginal gains from having access to additional working capital $W_i$:

$$(\varphi_i^K(\nu))^{-1} - 1 = \frac{d\pi_i(\nu)}{dW_i(\nu)}.$$ 

The variable profit earned by firm $\nu$ can be found by subtracting variable costs from revenue:

$$\pi_i(\nu) = p_i q_i(\nu) - w \ell_i(\nu) - \sum_{j=1}^{S} p_j m_{ij}(\nu).$$

For labor and unconstrained intermediate inputs, the lack of a wedge (that differs from one) in (1.8) and (1.9) implies that the private return to spending on these inputs is 1 and the marginal product of these inputs is equal to their marginal costs:

$$PR_i^l(\nu) = PR_i^j(\nu) = 1 \text{ for } j \notin K_i.$$  \hspace{1cm} (1.11)

To simplify notations, in what follows I let $\alpha_i(\nu)$ denote firm $\nu$'s equilibrium output elasticity with respect to labor, which is also equal to the labor expenditure share because labor is not subject to the credit constraint. Let $\sigma_{ij}(\nu)$ and $\omega_{ij}(\nu)$ respectively denote the output elasticity
and (potentially distorted) expenditure share of intermediate input $j$:

$$\alpha_i(\nu) \equiv \frac{\partial \ln f_i(\nu)}{\partial \ln \ell} = \frac{w\ell_i(\nu)}{p_iq_i(\nu)}, \quad \sigma_{ij}(\nu) \equiv \frac{\partial \ln f_i(\nu)}{\partial \ln m_{ij}}, \quad \omega_{ij}(\nu) \equiv \frac{p_jm_{ij}(\nu)}{p_jq_i(\nu)}.$$  

**Sectoral Allocations**  Let $N_i$ be the number of firms that enter sector $i$ in equilibrium. Recall sectoral total output and inputs are defined as

$$Q_i = N_i\mathbb{E}_\nu [q_i(\nu)], \quad M_{ij} = N_i\mathbb{E}_\nu [m_{ij}(\nu)], \quad L_i = N_i\mathbb{E}_\nu [\ell_i(\nu)],$$

where I use $\mathbb{E}_\nu [\cdot]$ to replace $\int \nu \cdot d\Phi_i(\nu)$ in Definition 1.1. The sectoral total expenditure on labor as a share of sectoral revenue, which I denote as $\alpha_i \equiv \frac{wL_i}{p_iQ_i}$ (without the index for firms, $\nu$), can be expressed as a weighted average of firm-level labor share, with weights being each firm’s output:

$$\alpha_i \equiv \frac{wL_i}{p_iQ_i} = \mathbb{E}_\nu \left[\frac{w\ell_i(\nu)}{p_iq_i(\nu)} q_i(\nu) \right] = \mathbb{E}_\nu \left[\frac{w\ell_i(\nu)}{p_iq_i(\nu)} \frac{q_i(\nu)}{\mathbb{E}_\nu [q_i(\nu)]} \right] $$

The sectoral expenditure share of intermediate inputs, which I denote as $\omega_{ij} \equiv \frac{p_jM_{ij}}{p_jQ_i}$, can be similarly expressed as

$$\omega_{ij} \equiv \frac{p_jM_{ij}}{p_jQ_i} = \mathbb{E}_\nu \left[\frac{p_jM_{ij}(\nu)}{p_jQ_i} \omega_{ij}(\nu) \right].$$  

The number of firms $N_i$ is pinned down by the free-entry condition (1.2), which can be expressed as

$$\frac{\kappa N_i}{p_iQ_i} = 1 - \alpha_i - \sum_{j=1}^{S} \omega_{ij}. \quad (1.13)$$

**Equilibrium**  To characterize the equilibrium, I take note that despite firm’s production functions being convex-concave, the economy features sectoral and aggregate constant returns to scale. This is because I allow for the entry of ex-ante identical entrepreneurs into sectors that produce homogeneous goods: any firm-level profits induced by concavity will be driven down to zero in net of fixed cost, and as a result the number of firms can be viewed as a flexible input at the sector level. Taking out input-output linkages, my sectoral production model is indeed a static version of the dynamic competitive model with entry studied by Hopenhayn (1992), Hopenhayn and Rogerson (1993), and more recently, by Restuccia and Rogerson (2008) and Buera et al. (2011, 2015).

Given input prices $(w, \{p_j\})$, the cost of producing $q$ units of output can be captured by the sectoral cost function, which is the solution to the dual of the entry and profit maximization
problem:

\[
TC_i (q; w, \{pj\}) \equiv \min_{n, \{\ell_i(\nu), \{m_{ij}(\nu)\}_j\}} n \left( n + \int_{\nu} \left( w\ell_i (\nu) + \sum_{j=1}^{S} p_j m_{ij}(\nu) \right) d\Phi_i (\nu) \right)
\]

s.t. \( n \int_{\nu} z_i (\nu) q_i (\nu) d\Phi_i (\nu) \geq q \)

\[
\sum_{j \in K_i} p_j m_{ij} \leq W_i (\nu)
\]

\[
q_i (\nu) = f_i (\ell_i (\nu), m_{i1}(\nu), \ldots, m_{iS}(\nu))
\]

A direct implication of the constant returns to scale property is that the sectoral cost function is linear in the level of output \( q \). In other words, the sectoral unit cost of production, which I write as

\[
C_i (w, \{pj\}) \equiv \frac{TC_i (q; w, \{pj\})}{q}
\]

is a function of only input prices but not output levels. Moreover, because the production function \( F(\cdot) \) of the final good also features constant returns to scale, I can write its unit cost function as

\[
C_F (\{pj\}) \equiv \min_{\{\bar{Y}_j\}} \sum_j p_j \bar{Y}_j \quad \text{s.t.} \quad F (\bar{Y}_j) \geq 1
\]

Equilibrium prices \( (w, \{pj\}) \) solve the set of equations

\[
C_i (w, \{pj\}) = p_i \quad \text{for all} \ i
\]

\[
C_F (\{pj\}) = 1
\]

where (1.17) reflects the normalization that the price of the final good is 1.

**Proposition 1.1.** There exists a unique decentralized equilibrium.

The intuition for the result is as follows. In this economy, the set of prices \( (w, \{pj\}) \) completely pins down equilibrium allocations. First, firm-level allocations are directly pinned down by prices, and solving for equilibrium boils down to deriving total sectoral level of output and inputs. Second, because labor supply is exogenous, the wage rate pins down aggregate consumption as \( C = wL \). Third, given that sectoral production features constant returns to scale, when holding input prices constant, the expenditure share on inputs is constant for both intermediate producers and the final producer. Given the level of aggregate consumption, we know the quantities of intermediate goods that go into the production of aggregate consumption. Next, given intermediate expenditure shares, we know the second round quantities of intermediate goods as well as the number of firms in each sector that go into the production of the intermediate goods that are used directly for the production of aggregate consumption. Iterating this logic ad infinitum and sum over quantities of
goods at each iteration, we can derive the total input and output levels in each sector. The infinite sum is well defined because labor share is positive in every industry by Assumption 1.1, and as a result, intermediate shares sum to less than one.

The uniqueness of the decentralized equilibrium therefore depends on the uniqueness of the price vector that satisfies the unit cost equations (1.16) and (1.17). My model is nested under the class of generalized Leontief models, and the standard argument of uniqueness for this class of models also applies to this setting (e.g., see Stiglitz 1970, Arrow and Hahn 1971), in which the Jacobian matrix of the mapping that represents the system of unit cost equations has the dominant diagonal property, which ensures the global uniqueness of solution by the classic results of Gale and Nikaido (1960).

1.2.3 Influence and Sales

We now proceed to better understand how credit constraints affect equilibrium allocations and sectoral sales. Recall \( \omega_{ij} \) denotes sector \( i \)'s expenditure share on intermediate good \( j \) and, as shown in equation (1.12), can be re-written as a weighted average of firm-level expenditure shares with weights being each firm’s output. I define a similar object based on firm-level elasticities:

\[
\sigma_{ij} = \mathbb{E}_\nu \left[ \sigma_{ij} (\nu) \frac{q_i (\nu)}{\mathbb{E}_\nu [q_i (\nu)]} \right].
\]

That is, \( \sigma_{ij} \) captures the proportional change in total output of sector \( i \) if every firm in sector \( i \) expands its use of intermediate input \( j \) by 1%. I refer to \( \sigma_{ij} \) as the sectoral output elasticity with respect to input \( j \), and indeed it is the elasticity of sectoral unit cost with respect to the price of inputs:

**Proposition 1.2.** In equilibrium,

\[
\sigma_{ij} = \frac{\partial \ln C_i (w, \{p_j\})}{\partial \ln p_j} \quad \text{for all } i, j.
\]

Absent credit constraints, \( \sigma_{ij} (\nu) = \omega_{ij} (\nu) \) for all firms, and as a result, sectoral expenditure shares are also equal to sectoral elasticities. On the other hand, when the constraints bind for a positive measure of firms, the sectoral expenditure shares of constrained inputs are distorted downwards relative to the equilibrium elasticities, with \( \omega_{ij} < \sigma_{ij} \) for \( j \in K_i \). Furthermore, the presence of additional constraints in the cost minimization problem (1.14) implies that if input prices are held fixed, the unit cost of output is higher when firms in a sector are subject to constraints.

I define the sectoral wedge on input \( j \) as the ratio between sectoral expenditure share and average
sectoral elasticity:

\[ \varphi_{ij} \equiv \frac{\omega_{ij}}{\sigma_{ij}} \leq 1. \]

If every firm in sector \( i \) expands its use of constrained input \( j \) by 1%, the total increase in sectoral sales would be \( \sigma_{ij} \% \) while the cost of using these additional inputs is \( \omega_{ij} \% \) of the sectoral sales. The inverse sectoral wedge on input \( j \), \( (\varphi_{ij})^{-1} \), can therefore be interpreted as the average private return to expenditure on input \( j \) as it captures the ratio between the marginal product and marginal cost of an uniform expansion in the use of input \( j \) across all firms in the sector. It can also be written as the average firm-level private returns to expenditure on constrained inputs, \( \varphi^K_i(\nu) \) (c.f. Lemma 1.1), weighted by the level of good \( j \) used by each firm:

\[ PR^j_i \equiv (\varphi_{ij})^{-1} = \mathbb{E}_\nu \left[ \left( \varphi^K_i(\nu) \right)^{-1} \frac{m_{ij}(\nu)}{\mathbb{E}_\nu [m_{ij}(\nu)]} \right] \quad \text{for } j \in K_i. \]

Relatedly, I denote the sectoral average private return to capital inputs as a whole by

\[ PR^K_i \equiv (\bar{\varphi}_i^K)^{-1} = \frac{\sum_{j \in K_i} \sigma_{ij}}{\sum_{j \in K_i} \omega_{ij}} \mathbb{E}_\nu \left[ \left( \varphi^K_i(\nu) \right)^{-1} \frac{\sum_{j \in K_i} p_j m_{ij}(\nu)}{\mathbb{E}_\nu \left[ \sum_{j \in K_i} p_j m_{ij}(\nu) \right]} \right]. \] (1.18)

When the production function \( f_i(\cdot) \) is homothetic, \( \varphi_{ij} = \varphi_i^K \) for all \( j \in K_i \).

The sectoral average private return of unconstrained inputs, including labor and intermediate inputs \( j \notin K_i \), is equal to one, just as the firm-level counterparts in equation (1.11):

\[ PR^j_i = PR^j_i = 1 \quad \text{for } j \notin K_i. \]

Because a smaller fraction of revenue is spent on inputs, a larger fraction must accrue to variable profits and attract firm entry. Indeed, equation (1.13) reveals that, holding input prices fixed, when firms in a sector are constrained, more firms enter per unit of sectoral output relative to when there are no constraints in the sector. Intuitively, financial constraints manifest themselves at the sector level by creating wedges between the marginal product and marginal cost for both the constrained intermediate inputs and the number of firms that are established. When a sector is constrained, resources are misallocated within the sector, with too many firms in equilibrium, each using too little of the constrained inputs. While the result may seem counterintuitive, it is merely a statement about a local property of the equilibrium and does not imply that discrete changes to the environment, such as removing credit constraints altogether from a sector, would induce fewer firms to be in the new equilibrium. Furthermore, the result is not necessarily at odds with empirical observations: Hsieh and Olken (2014) find that the distributions of manufacturing firm size in India, Indonesia, and Mexico are skewed to the left relative to that in the U.S., with
the developing countries having many more small firms relative to medium and large firms.

I now define three important objects that are central to the analysis. Recall that $h_i$ denotes the Hicks-neutral sectoral productivity that is common to all firms in sector $i$, which is normalized to 1 throughout the exposition. The notation is introduced solely for the purpose of the following definition:

**Definition 1.3.** The influence vector $\mu' \equiv (\mu_1, \ldots, \mu_S)$ is the elasticity of net aggregate output $Y$ with respect to sector productivity,

$$\mu_i \equiv \frac{d \ln Y}{d \ln h_i}.$$  

**Definition 1.4.** The sales vector $\gamma' \equiv (\gamma_1, \ldots, \gamma_S)$ is the ratio between total sectoral sales and net aggregate output,

$$\gamma_i \equiv \frac{p_i Q_i}{Y}.$$  

**Definition 1.5.** The sales gap vector $\xi' \equiv (\xi_1, \ldots, \xi_S)$ is the element-wise ratio between influence and sales:

$$\xi_i \equiv \frac{\mu_i}{\gamma_i}.$$  

The influence vector $\mu'$ is a notion of sectoral importance, whereas the sales vector $\gamma'$ represents the equilibrium size of sectors. The sales gap $\xi_i$ captures the wedge between sectoral importance and size and is a key object in the policy analysis. To better understand these objects and how credit constraints endogenously affect them, I first go to the specific example $E$.

**Influence and Sales in Example $E$** The expenditure shares on intermediate goods in sectors 2 and 3 are:

$$\frac{p_1 M_{21}}{p_2 Q_2} = \sigma_2 \varphi_2^K, \quad \frac{p_2 M_{32}}{p_3 Q_3} = \sigma_3 \varphi_3^K,$$

where $\sigma_i$ is the output elasticity in sector $i$ with respect to intermediate input and $\sigma_i \varphi_i$ is sector $i$'s expenditure share on the constrained intermediate input, with $\varphi_i < 1$ iff the constraint in sector $i$ binds.

In this example, the influence, sales, and sales gap are respectively:

$$\mu' \propto (\sigma_3 \sigma_2, \sigma_3, 1),$$

$$\gamma' \propto \left( (\sigma_3 \varphi_3^K), (\sigma_2 \varphi_2^K), \sigma_3 \cdot \varphi_3^K, 1 \right),$$

$$\xi' \propto \left( (\varphi_3^K \varphi_2^K)^{-1}, (\varphi_3^K)^{-1}, 1 \right).$$

31
I highlight three observations. First, the influence of downstream sector 3 is larger than that of midstream sector 2, which in turn has larger influence than upstream sector 1. This is because the final good is produced directly from the downstream good, and any productivity shock in sector 3 will directly affect the effective aggregate productivity, whereas positive productivity shocks in up- and midstream sectors will affect the effective aggregate productivity only through their indirect effect on the relative price of good 3. A similar intuition applies to more general network structures: the sectors with high influence will be those that heavily supply to the final good either directly or indirectly through other sectors.

The second observation relates to how credit wedges $\varphi^K_i$ affect sales and the sales gap. In this economy, the entire output of sector 2 is used as inputs by sector 3, hence the total sales of sector 2 relative to those of sector 3 is captured by $\sigma_3 \varphi^K_3$, the intermediate expenditure share of sector 3. Similarly, sales of sector 1 relative to sector 2 is simply $\sigma_2 \varphi^K_2$. Note that sector 2’s relative sales are affected by $\varphi^K_3$ but not $\varphi^K_2$: in other words, it is the credit constraints faced by downstream buyers, not within the sector itself, that affect the relative size of sector 2. Furthermore, sales is most suppressed in upstream sector 1, despite the fact that sector 1 itself is unconstrained. This is because sector 1’s size is affected by constraints in both midstream and downstream sectors—an effect that is multiplicative in the sectoral wedges. The further upstream we go, and as we travel through an increasing number of constrained sectors, the higher sales gap we would find of a sector. In equilibrium, the upstream sectors are too small in sales relative to their influence, while the downstream ones are too large.

Lastly, absent credit constraints, $\varphi^K_2 = \varphi^K_3 = 1$ and influence equals sales. This property holds under my general model and is originally formalized by Hulten (1978).

**Influence and Sales in the General Model** I now proceed to derive influence and sales in the general model and extend the intuitions to this environment. To find influence and sales in equilibrium, it is convenient to stack the sectoral elasticities and expenditure shares into matrices $\Sigma$ and $\Omega$:

$$
\Sigma = \begin{bmatrix}
\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\
\sigma_{21} & \sigma_{22} & \cdots & \sigma_{2M} \\
\vdots & \ddots & \ddots & \vdots \\
\sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM}
\end{bmatrix},
\Omega = \begin{bmatrix}
\omega_{11} & \omega_{12} & \cdots & \omega_{1M} \\
\omega_{21} & \omega_{22} & \cdots & \omega_{2M} \\
\vdots & \ddots & \ddots & \vdots \\
\omega_{M1} & \omega_{M2} & \cdots & \omega_{MM}
\end{bmatrix}.
$$

Each row in the matrices represents an output sector, while each column represents an input sector. That is, $\sigma_{ij}$, or the entries on the $i$-th row and $j$-th column of the matrix $\Sigma$, represents the sectoral output elasticity in sector $i$ of input $j$. Similarly, $\omega_{ij}$ is the share of expenditure on input $j$ as a fraction of the total sales in sector $i$. For this reason, $\Omega$ represents the input-output table of the economy and is directly observable from national accounts. In an economy without any distortions
or tax interventions, the output elasticity matrix coincides with the input-output table: $\Omega \equiv \Sigma$. The entries $\sigma_{ij}$ and $\omega_{ij}$ differ precisely because of sectoral distortions such as financial constraints and tax interventions.

Let $\beta'$ denote the equilibrium vector of expenditure share of the final good producer,

$$\beta' \equiv \left( \frac{p_1 Y_1}{F(Y_1, \ldots, Y_S)}, \ldots, \frac{p_S Y_S}{F(Y_1, \ldots, Y_S)} \right),$$

which is referred to as the vector of final shares. Because the final producer is unconstrained, $\beta'$ also represents the equilibrium vector of final good's output elasticities with respect to inputs, i.e. $\beta_i = \frac{\partial \ln F(Y_1, \ldots, Y_S)}{\partial \ln Y_i}$.

**Proposition 1.3.** In the decentralized equilibrium,

a) The influence vector $\mu'$ equals

$$\mu' = \frac{\beta' (I - \Sigma)^{-1}}{\beta' (I - \Sigma)^{-1} \cdot \alpha'},$$

where $\alpha'$ is the vector of sectoral output elasticity with respect to labor.

b) The sales vector $\gamma'$ equals

$$\gamma' = \frac{\beta' (I - \Omega)^{-1}}{\beta' (I - \Omega)^{-1} \cdot \alpha'}.$$

**Corollary 1.1.** (Hulten 1978) Absent credit constraints, $\omega_{ij} = \sigma_{ij}$ for all $i, j$, and influence equals to sales.

The fact that influence is equal to sales absent market imperfections is first shown by Hulten (1978) on the class of generalized Leontief models with aggregate constant returns to scale, and it is the basis for using sales to measure sectoral importance in the growth accounting literature. This equivalence holds in my model when there are no credit constraints but is otherwise broken. I now provide the intuition for why this is the case through the lens of the general model.

The object $(I - \Sigma)^{-1} = I + \Sigma + \Sigma^2 + \cdots$ is the Leontief inverse of the sectoral output-elasticity matrix $\Sigma$. This object, important in the input-output literature, summarizes how sectoral productivity shocks propagate downstream to other sectors through the infinite hierarchy of cross-sectoral linkages. To understand why influence takes the form in the proposition, consider normalizing all prices by wage rate and hold constant the fixed cost of entry relative to the wage rate at $\gamma/w$. An one-percent increase in Hicks-neutral productivity $h_j$ in sector $j$ has the direct effect of lowering output prices in its downstream sector $i$ by $\sigma_{ij}$ percent, represented by the $ij$-th entry of the output elasticity matrix $\Sigma$. The shock also has a second order effect that lowers output prices for all goods $k$ that use $j$'s output as inputs, which in turn further lowers the prices in sector $i$. This second order effect is captured by the $ij$-th entry of the matrix $\Sigma^2$, and so on. The $ij$-th entry in the Leontief inverse matrix $[(I - \Sigma)^{-1}]_{ij}$ therefore captures the total effect of a productivity shock in
sector $j$ on the output price of sector $i$. These effects then translate into higher aggregate output (or equivalently, lower relative price of the final good to wage rate), reflected by the dot product between the output elasticity of the final good $\beta'$ and the Leontief inverse $(I - \Sigma)^{-1}$. In sum, the sectors with high influence in an economy are those with high network-adjusted final shares.

The scalar term $\frac{1}{\beta'(I - \Sigma)^{-1}\beta'}$, which is not present in formulations like Acemoglu et al. (2012), arises from the endogenous entry of firms in my model. As the final good becomes cheaper relative to the wage rate, entry becomes less costly. This attracts to more firms to enter all industries, creating an amplification effect.

The sales vector takes the same form as the influence vector, replacing the elasticity matrix $\Sigma$ with the expenditure share matrix $\Omega$. To see why sales are constructed with the expenditure share matrix, note that sectoral sales can be written as the infinite sum that consists of 1) its output supplied to produce the final good; 2) its output used by other sectors to produce the final good; 3) its output used by other sectors, which supply to other sectors to produce the final good, and so on:

$$p_j Q_j = p_j C_j + \sum_{i=1}^{S} p_i C_i \omega_{ij} + \sum_{i=1}^{S} p_i C_i [\Omega^2]_{ij} + \cdots$$

The common denominator in the sales vector reflects the fact that only a fraction of the final good accrues to net aggregate output while the remaining fraction is used to incur the overhead fixed cost of entry.

There are two ways in which financial frictions affect equilibrium allocations. First, as discussed earlier, constraints within a sector lower the effective sectoral productivity by increasing the price of sectoral output when input prices are held constant. This effect travels downstream, serving as a negative productivity shock that increases the price of all downstream goods and eventually the final good.

Second and more central to my analysis, financial frictions also affect the relative sectoral size and distort sales away from influence. Credit constraints suppress equilibrium demand of constrained intermediate inputs, endogenously affecting the equilibrium input-output linkages by reducing the sales of upstream goods that are subject to constraints. Contrary to the downstream travel of productivity shocks, this effect instead travels upstream, as can be seen from equation (1.20). Credit constraints in sector $i$ reduce the equilibrium sales of good $j$ to sector $i$. Even if sector $j$ is not constrained, the sector still uses fewer inputs from its own upstream suppliers because it faces less demand for its output, and in turn these upstream suppliers end up with lower sales. In equilibrium, it is the sectors that supply to many constrained sectors, which in turn supply to many constrained sectors, ad infinitum, that have the least equilibrium sales relative to influence.

Baqae (2015) and Acemoglu et al. (2016) observe that in a production network under Cobb-Douglas technology assumption, productivity shocks travel downstream through input-output linkages from suppliers to buyers, while demand shocks travel upstream. My analysis so far makes two
additional contributions to understanding how shocks propagate. First, financial frictions serve both as a productivity shock and a shock to intermediate demand that emanates from the constrained sectors. The productivity shock aspect of financial frictions propagates downstream by lowering aggregate output, while the demand shock aspect propagates upstream and suppresses the relative size of upstream sectors. Second, my analysis shows that these results do not rely on specific functional form assumptions. Demand shocks have no effect on the prices of output or the unit costs of production in a generalized Leontief model with sectoral constant-returns-to-scale and affect equilibrium quantities only through backward linkages or, in other words, by traveling upstream. On the other hand, productivity shocks travel only through forward linkages, affecting the unit cost of production and equilibrium prices of downstream buyers. Equilibrium prices of upstream sectors are unaffected, and output quantities in upstream sectors change in response to productivity shocks from downstream only through the changes in demand induced by these shocks.

1.2.4 Industrial Policies

I now proceed to show that the sales gap, i.e. the ratio between sectoral influence and sales, is a sufficient statistic that could guide policy. There is clearly room for policy intervention in this economy. Even without relaxing credit constraints, if a planner could impose firm-level subsidies and taxes on production inputs and profits, first-best allocations can be restored. Specifically, the planner would tax firm profits to reduce entry in constrained sectors while imposing firm-specific subsidies to constrained inputs and undo the wedges imposed by credit constraints. The level of firm-specific subsidies that restore first-best would be such that the Lagrange multiplier on credit constraints is precisely zero. However, to implement such policies successfully in any real-world economies, a benevolent government has to grapple with two difficulties. First, the planner needs to have the fiscal flexibility to tailor subsidies to individual firms within sectors. Second, the planner has to know the exact nature of credit constraints for each firm, which requires not only information on each firm’s amount of working capital and trade credit but also knowledge of firm-level productivities. The required information and fiscal flexibility in first-best implementation are luxuries that most policymakers do not have. For this reason, I consider tax instruments that apply to all firms equally within a given sector.

To make progress, I leverage one crucial feature of generalized Leontief models: not only do distortions generated by credit constraints propagate through input-output linkages, but so does the effect of policy interventions. Due to pecuniary externalities from the input-output linkages, subsidizing production upstream lowers the prices of upstream goods, which indirectly relaxes credit constraints downstream and ameliorates cross-sector resource misallocation. This property of production networks leaves room for welfare-improving policy interventions, even when the planner only has access to a limited set of instruments. My main results show that the sales gap exactly

---

1This result can be viewed as an application of the famous non-substitution theorem by Samuelson (1951).
captures the ratio between the social and private marginal return of spending resources on sectoral production, starting with the decentralized no-tax equilibrium. Rather than making assumptions about the set of instruments at the planner’s disposal and prescribing optimal policies under these assumptions, an exercise I conduct under Cobb-Douglas assumptions, these results instead provide answers to the following question: starting from the decentralized equilibrium without any government intervention, where should the fiscal authority spend the first dollar of its tax budget, among a given set of linear instruments that induce firms to use more production inputs? The result is especially usable because the theory makes no assumptions about the availability and flexibility of tax instruments: policymakers can use the theory to compute social returns given the respective fiscal constraints they face.

**Equilibrium with Taxes** The planner is has access to lump-sum tax $T$ on the representative consumer’s wage income and a flexible set of linear subsidies $\{\tau_i^R, \tau_i^L, \tau_i^1, \ldots, \tau_i^S\}_{i=1}^S$ that applies to either sales or production inputs in sector $i$. The planner also has some real, non-tax expenditure $E$ that is financed by lump-sum tax $T$. This could capture expenditures on public goods or other forms of public consumption. I introduce $E$ merely for interpretational purposes, and do not explicitly model how the planner and the representative consumer value $E$. In the presence of sectoral subsidies, the planner has to finance both the real expenditure $E$ and the subsidies by the lump-sum tax, with a budget constraint

$$T = E + \sum_{i=1}^S \left( \tau_i^R p_i Q_i + \tau_i^L w L_i + \sum_{j=1}^S \tau_i^j p_j M_{ij} \right). \quad (1.21)$$

The budget constraint of the representative consumer is

$$w L = C + T. \quad (1.22)$$

The resource constraint of the economy is

$$Y = C + E. \quad (1.23)$$

In presence of the subsidies, the profit maximization problem for firms in sector $i$ becomes

$$\begin{align*}
\left( p_{\text{Tax}_{i, \text{firm}}} \right) \max_{\{m_{ij}\}_{i=1}^S, \ell} \left( 1 + \tau_i^R \right) p_i q_i (\nu, \ell, \{m_{ij}\}) - \frac{w}{1 + \tau_i^L} \ell - \sum_{j=1}^S \frac{p_j}{1 + \tau_i^j} m_{ij} \\
\text{s.t.} \quad \sum_{j \in K_i} \frac{p_j}{1 + \tau_i^j} m_{ij} \leq W_i (\nu)
\end{align*}$$

36
I modify the definition of an equilibrium to incorporate taxes.

**Definition 1.6.** An equilibrium with taxes is a collection of subsidies \( \{ \tau_i^R, \tau_i^I, \tau_i^J, \ldots, \tau_i^S \}_{i=1}^S \), prices \( \{ p_i \}_{i=1}^S \), wage rate \( w \), measure of firms \( \{ N_i \}_{i=1}^S \), firm-level allocations \( \{ \ell_i (\nu), m_{i1} (\nu), \ldots, m_{iS}, q_i (\nu) \}_{i=1}^S \), production inputs for the final good \( \{ Y_i \}_{i=1}^S \), net aggregate output \( Y \), aggregate labor supply \( L \), aggregate consumption \( C \), lump-sum tax \( T \), and non-tax fiscal expenditure \( E \) such that (i) firms in the intermediate sectors solve the constrained profit maximization problem (P \( P_{i,firm} \)); (ii) free-entry drives ex-ante profits to zero in all intermediate sectors according to (1.2); (iii) the final good producer solves (1.3); (iv) the budget constraint (1.21) for the planner and (1.22) for the representative hold; (v) all markets clear such that (1.4), (1.5), (1.6), and (1.23) hold.

**Elasticity of Net Aggregate Output to Subsidies** All equilibrium allocations and prices can be written as functions of the exogenous subsidy vector \( \tilde{\tau} \), lump-sum tax \( T \), and expenditure \( E \). In the following exercise, I start from the no-subsidy equilibrium with \( \tilde{\tau} = 0 \) and some expenditure level \( E = T \) balanced by lump-sum tax. I evaluate the change in net aggregate output in response to a marginal increase in a subsidy for a given sector, balanced by a marginal increase in \( T \) while holding \( E \) constant.

**Theorem 1.1.** Starting from a decentralized equilibrium with no subsidies (\( \tilde{\tau} = 0 \)) and holding \( E \) constant,

a) The elasticity of net aggregate output with respect to labor subsidy in sector \( i \) is

\[
\frac{d \ln Y}{d \ln (1 + \tau_i^{L})} \bigg|_{\tilde{\tau} = 0, \text{holding } E \text{ constant}} = \alpha_i (\mu_i - \gamma_i).
\]

b) The elasticity of net aggregate output with respect to subsidy to unconstrained input \( j \in X_i \) in sector \( i \) is

\[
\frac{d \ln Y}{d \ln (1 + \tau_{ij})} \bigg|_{\tilde{\tau} = 0, \text{holding } E \text{ constant}} = \sigma_{ij} (\mu_i - \gamma_i).
\]

c) The elasticity of net aggregate output with respect to subsidy to constrained input \( j \in K_i \) in sector \( i \) is

\[
\frac{d \ln Y}{d \ln (1 + \tau_{ij})} \bigg|_{\tilde{\tau} = 0, \text{holding } E \text{ constant}} = \sigma_{ij} \mu_i - \omega_{ij} \gamma_i.
\]

d) The elasticity of net aggregate output with respect to sales (revenue) subsidy in sector \( i \) is

\[
\frac{d \ln Y}{d \ln (1 + \tau_i^{R})} \bigg|_{\tilde{\tau} = 0, \text{holding } E \text{ constant}} = \mu_i - \gamma_i.
\]
While the relative sectoral size can be captured by sales under equilibrium production, it is captured by influence under optimal production. Theorem 1.1 shows that the distance between influence and sales provides a direction in which production efficiency can be improved. To understand these results, consider the effect of a labor subsidy in sector $i$. Recall that $\alpha_i$ is the labor elasticity, and $\alpha_i \mu_i$ captures the proportional effect on aggregate output if every firm in sector $i$ expands its labor input by 1%, holding labor allocation in every other sector fixed. On the other hand, this exercise violates the resource constraint because the total labor endowment is fixed, and labor has to be scaled back from other sectors in order for $L_i$ to increase. By financing the labor subsidy via the lump-sum tax, labor scales back uniformly across all sectors, including $i$. The amount of labor that must be scaled back from every sector in order to balance the 1% increase in sector $i$ is captured by the total amount of labor hired by sector $i$ relative to the rest of the economy, i.e. the product between sector $i$'s sales and its labor share, $\alpha_i \gamma_i$. This product in turn captures the negative effect on net aggregate output $Y$ from scaling back labor, and the difference between the two terms is the total effect.

This intuition ignores the fact that resources reallocate endogenously in response to changes in labor allocation, but these re-allocative effects cancel out and have no additional impact on net aggregate output. This is due to the aggregation of the envelope conditions from each firm's optimization problem: although firms face additional credit constraints in their profit maximization problems, their optimization over constrained inputs ensures that envelope condition applies. The intuition is similar on the results for the other subsidies.

**Social Return to Tax Dollar Spent on Inputs** Consider again starting from the no-subsidy equilibrium with $\tau = 0$ and some fiscal expenditure $E = T$ balanced by lump-sum tax. Suppose the planner wants to implement a marginal subsidy $\tau_i^L$ to labor in sector $i$ but cannot raise any additional lump-sum tax and must balance the budget by cutting back on fiscal expenditure $E$. The following object captures the marginal change in total private consumption as a result of cutting back one dollar of fiscal expenditure $E$ and spending it by increasing $\tau_i^L$:

$$SR_i^L \equiv - \left. \frac{dC}{d\tau_i^L} \right|_{\tau=0, \text{holding } T \text{ constant}}$$

I refer to this object as the social return to expenditure on labor in sector $i$. I define the social return to expenditure on intermediate inputs by replacing $\tau_i^L$ in equation (1.24) with $\tau_i^I$:

$$SR_i^I \equiv - \left. \frac{dC}{d\tau_i^I} \right|_{\tau=0, \text{holding } T \text{ constant}}$$
The social return to sectoral expenditure on all inputs can be defined similarly by replacing $\tau_i^L$ with $\tau_i^R$, the subsidy to revenue or sales, which affects all inputs uniformly:

$$SR_i^R \equiv \frac{dC / d\tau_i^R}{dE / d\tau_i^R}$$

As the next theorem demonstrates, the social return to expenditures on inputs in a sector is closely related to the sales gap.

**Theorem 1.2.** Starting from a decentralized equilibrium with no subsidies ($\tau = 0$) and holding $T$ constant,

a) The social return to expenditure on labor in sector $i$ is

$$SR_i^L = \xi_i.$$

b) The social return to expenditure on unconstrained input $j \in X_i$ in sector $i$ is

$$SR_i^j = \xi_i \text{ for } j \in X_i.$$

c) The social return to expenditure on constrained input $j \in K_i$ in sector $i$ is

$$SR_i^j = (\varphi_{ij})^{-1} \xi_i \text{ for } j \in K_i.$$

d) The social return to a revenue subsidy in sector $i$ is

$$SR_i^R = \xi_i.$$

Recall that $(\varphi_{ij})^{-1}$ can be interpreted as the average sectoral private return to expenditures on constrained input $j$, which captures the ratio between the marginal product and marginal cost if every firm in sector $i$ expands its use of input $j$ uniformly. Theorem 1.2 can be restated as:

**Corollary 1.2.** The ratio between social and average private return to sectoral expenditure on inputs satisfies:

$$\frac{SR_i^L}{PR_i^L} = \frac{SR_i^j}{PR_i^j} \text{ (for all } j) = \xi_i.$$

The sales gap is a sufficient statistic that captures the ratio between social and private marginal return to expanding the use of production inputs in the sector. The result is especially usable because it answers the following question: if the planner has one dollar of tax budget to spare on providing subsidies to inputs, to which sector and to which input should the planner direct the subsidy?
The intuition behind Theorem 1.2 and the corollary is similar to that of Theorem 1.1, in that sectoral size scales with influence under optimal production and with sales under equilibrium production. In a sense, influence locally represents the potential sales vector that would have prevailed had there been no credit constraints. The ratio between the potential and actual sales summarizes the inefficiencies in the production network, and as long as we know the sales gap, knowledge of the underlying frictions in the input-output system becomes irrelevant for welfare analysis.

It is worth emphasizing that sectors with the highest sales gaps are not necessarily sectors in which firms are most constrained; instead, they are sectors that directly or indirectly supply to many constrained sectors. While the most constrained sectors have the highest private return to expenditure on capital goods, the social return might not be high in these sectors.

There are two additional and complementary intuitions through the three-sector example economy $E$ for why input subsidies applied to higher-sales-gap sectors provide higher social returns to tax dollars. Recall that in the example, the vector of sales gaps according to Theorem 1.2 is

\[ (\xi_1, \xi_2, \xi_3) \propto \left( \frac{1}{\varphi_1^K \varphi_2^K}, \frac{1}{\varphi_3^K}, 1 \right). \]

The first intuition is that there are prices in credit constraints, and hence pecuniary externalities do not net out in this economy (Greenwald and Stiglitz 1986). In fact, suppliers' prices are what show up in buyers' constraints. Subsidizing production for upstream suppliers indirectly relaxes the credit constraints of midstream buyers, who are then able to expand production and further relax the constraints of downstream producers. The higher a sector's sales gap, the greater is this effect.

The second intuition relates directly to the allocation of productive resources. Credit constraints from downstream firms reduce demand for goods produced upstream, which in turn reduces the amount of inputs allocated to upstream sectors. The higher a sector's sales gap, the more severe is the misallocation of production input to that sector due to the cumulative rounds of distortions as goods change hands from up- to downstream firms. Corollary 1.2 points out that the ratio between social and private return to expenditure on inputs is exactly captured by the degree to which sectoral inputs are misallocated due to frictions along input-output linkages.

In Appendix 1.9, I show that Corollary 1.2 holds true even when policy instruments target only a subset of firms (rather than all firms) within each sector: that is, the sales gap still captures the ratio between social and private marginal returns of tax expenditure for these policy instruments.

**Directed Credit** Consider a modified economic environment in which credit constraints can be relaxed at a cost. That is, suppose the planner controls an instrument $\tau_i^C$ that relaxes the credit constraints faced by firms in sector $i$ according to

\[ \sum_{j \in K_i} p_j m_{ij} \leq (1 + \tau_i^C) W_i (\nu). \] (1.25)
A firm’s problem under the additional working capital is to solve

$$\max_{\{m_{ij}\}_{i=1}^S, \ell} p_i q_i (\nu, \ell, \{m_{ij}\}) - \sum_{j=1}^S p_j m_{ij} - w \ell$$

subject to (1.25).

In order to deliver any additional credit, the planner has to incur a monitoring cost (in terms of the final good) that is linearly proportional to the amount of additional credit taken up by constrained producers, which can be expressed as

$$DC \left( \left\{ \tau_i^C \right\} \right) = \chi \sum_i N_i \left( \left\{ \tau_i^C \right\} \right) E_{\nu} \left[ \max \left\{ \left( \sum_{j \in K_i} p_j m_{ij} (\nu) \right) - W_i (\nu), 0 \right\} \right], \quad (1.26)$$

where the max operator reflects the fact that not all firms are constrained and the linear monitoring cost $\chi$ is only incurred on the portion of additional working capital that prevails in equilibrium due to the instruments $\left\{ \tau_i^C \right\}$. We modify the budget constraint for the planner as

$$T = E + DC. \quad (1.27)$$

To understand the effect of directed credit on equilibrium, consider a marginal increase $dW_i (\nu)$ in working capital available to firm $\nu$ starting from the decentralized equilibrium. The gain captured by the firm is

$$d\pi_i (\nu) = dW_i (\nu) \times (\varphi_i^K (\nu)^{-1} - 1)$$

and the cost of delivering the additional working capital is $1 \left( \varphi_i^K (\nu)^{-1} > 1 \right) \cdot \chi \cdot dW_i (\nu)$. Therefore, for a marginal change $d\tau_i^C$ which uniformly relaxes the credit constraints for all firms in sector $i$, the ratio between the gains captured by firms in sector $i$ and the total monitoring cost is

$$PR_i^C = \chi^{-1} \cdot E \left[ (\varphi_i^{-1} (\nu) - 1) \frac{W_i (\nu)}{E [W_i (\nu) 1 (\varphi_i (\nu)^{-1} > 1)]} \right]. \quad (1.28)$$
I refer to $PR_i^C$ as the private return to credit\footnote{This formulation of directed credit assumes that the monitoring cost is incurred only for the amount of additional working capital that is taken up by firms. The environment can be modified trivially to accommodate the alternative assumption that a cost of $rFW_i(\nu)$ has to be incurred regardless of whether the credit constraint binds for firm $\nu$: simply replace equation (1.26) with $DC\{\tau_i^C\} = \chi \sum_i N_i \left(\tau_i^C\right) E\left[\tau_i^C W_i(\nu)\right]$ and remove the indicator function $1(\varphi_i(\nu)^{-1} > 1)$ from equation (1.28). Proposition 1.4 below applies to this modified environment.}. On the other hand, the social return to credit is

$$SR_i^C = \frac{dC/d\tau_i^C}{dE/d\tau_i^C} \quad \text{if } \tau = 0, \text{holding } T \text{ constant}$$

$$= \frac{dC/d\tau_i^C}{dDC/d\tau_i^C} \quad \text{if } \tau = 0, \text{holding } T \text{ constant}$$

We have a result in this environment that is analogous to Corollary 1.2.

**Proposition 1.4.** The ratio between the social and private return to credit instrument $\tau_i^C$ is captured by the sales gap:

$$\frac{SR_i^C}{PR_i^C} = \xi_i.$$

Proposition 1.4 points to a powerful intuition: suppose the return to credit is equalized across all firms and all sectors, but firms are still constrained with $\varphi_i(\nu) = \varphi < 1$ for all $i$ and $\nu$. In this case, the social return to working capital would not be the same across sectors. Instead, the social return is highest precisely for the sectors that tend to be upstream and supply, directly or indirectly, to many constrained buyers\footnote{This case is analyzed in Appendix 1.8.1, in which I model financial frictions as a linear monitoring cost $\chi_i$ and endogenize lending by allowing firms to choose $W_i(\nu)$ freely and pay the monitoring cost as interests. In that environment, when $\chi_i \equiv \chi$ for all $i$, the marginal private return to credit is equalized across all sectors with $\varphi_i = \frac{1}{1+\chi}$ but the sales gap still captures the ratio between social and private marginal returns to credit and it differs from one in general as long as $\chi > 0.$}. This result offers a potential explanation of why industrial policies often direct credit to upstream sectors. More importantly, we see that it might be in the planner’s interest to impose restrictions on private credit markets because their operations could compromise the efficient social allocation of working capital.

The result in Proposition 1.4 assumes a proportional increase of working capital for all firms within a given sector, but a similar result can be obtained if instead the working capital relaxation is uniform in levels across firms. To see this, consider instrument $\tau_i^{C'}$ that relaxes the credit constraints according to

$$\sum_{j \in K_i} p_{ij} \leq W_i(\nu) + \tau_i^{C'}.$$ 

I again make the assumption that the cost of delivering the additional working is proportional to the actual amount taken up by firms as in equation (1.26). The private return for a marginal
change in $\tau_i^{C'}$ is

$$PR_i^{C'} = \frac{\mathbb{E} \left[ d\pi_i (v) / d\tau_i^{C'} \right]}{\chi \Pr (\varphi_i (v)^{-1} > 1)}$$

$$= \chi^{-1} \cdot \mathbb{E} \left[ (\varphi_i^{-1} (v) - 1) \left| \varphi_i (v)^{-1} > 1 \right. \right].$$

and the social return is again defined as

$$SR_i^{C'} = \frac{dC/d\tau_i^{C'}}{dE/d\tau_i^{C'}} \bigg| _{\tau=0, \text{holding } T \text{ constant}}.$$

**Proposition 1.5.** The ratio between the social and private return to credit instrument $\tau_i^{C'}$ is captured by the sales gap:

$$\frac{SR_i^{C'}}{PR_i^{C'}} = \xi_i.$$

**The Irrelevance of Social Welfare Functions** The welfare results in Theorem 1.2 and Proposition 1.4 provide guidance as to how the planner should spend the first dollar of the tax budget given a set of feasible instruments. In choosing which tax instrument to use and which sector to subsidize, how the planner marginally trades off between private and public consumption is irrelevant. To see this, let $U (C, E)$ denote the social welfare function. The marginal change in $U$ following an intervention that cuts back $E$ by one dollar to finance a subsidy $r_i$ is

$$\frac{dU/d\tau_i}{dE/d\tau_i} \bigg| _{\tau=0, \text{holding } T \text{ constant}} = \frac{\partial U}{\partial E} + \frac{\partial U / dC / d\tau_i}{\partial C / dE / d\tau_i} \bigg| _{\tau=0, \text{holding } T \text{ constant}}.$$

It is immediate apparent that the planner's preference ranking over subsidies is completely captured by the ranking of the social returns.

**Optimal Labor Subsidy under Cobb-Douglas Production Functions** My next result pertains to optimal (as opposed to marginal) linear subsidies to labor in production. Labor is special relative to other production inputs because it is the only exogenous factor and the only source of net value-added in this economy. With iso-elastic firm-level production functions or Cobb-Douglas sectoral technologies, both the elasticity matrix and the input-output table are stable. In this case, the sales gap captures not only the marginal social return to expanding labor inputs but also the subsidies at the social optimum if the planner can freely choose any level of $\tau_i^l$.

**Theorem 1.3.** Suppose all production functions are iso-elastic, with

$$f_i (v) = \ell (v)^{\alpha_i} \prod_{i=1}^{S} m_{ij} (v)^{\sigma_{ij}}$$
for firms and
\[ F(\{Y_i\}) = \prod_{i=1}^{S} Y_i^{\beta_i} \]
for the final good. The optimal value-added subsidies, i.e. the solution to the planning problem
\[ \tilde{\tau}_L = \arg \max_{\{\tau_i^L\}} Y \left( \{\tau_i^L\} \right) \]
satisfies
\[ 1 + \tau_i^L \propto \xi_i. \tag{1.29} \]

The result as stated is on the proportionality of \( 1 + \tau_i^L \) because the levels are not pinned down: having access to unrestricted usage of value-added tax is a substitute for lump-sum tax on consumers, or a uniform tax on wages—the planner can always scale \( 1 + \tau_i^L \) by a constant and adjust the lump-sum tax accordingly to balance the budget.

**Sales Gap and Hirschmanian Linkages** I conclude this theory section with a closed-form formula for the sales gap measure in the general model, and I use this result to place the measure in a historical context and connect it to an early literature that follows from the seminal work of Hirschman (1958).

As discussed, distortions in sales pass through from downstream to upstream sectors through backward demand linkages, and in the three-sector example \( \mathcal{E} \), the pass-through is complete: even if midstream sector 2 is unconstrained with \( \varphi_2^K = 1 \), the sales of upstream sector 1 are still distorted to exactly the same degree as those of sector 2 relative to their influence. This is because midstream is the sole buyer of upstream good in that example. Under more general network structures, the pass-through of sales gap from sector \( i \) (buyer) to sector \( j \) (seller) depends on the importance of \( i \) as a buyer of sector \( j \)'s output. To capture these notions, let
\[ \hat{\omega}_{ij} \equiv \frac{p_j M_{ij}}{p_j Q_j} = \omega_{ij} \frac{\gamma_i}{\gamma_j}, \quad \hat{\sigma}_{ij} \equiv \sigma_{ij} \frac{\gamma_i}{\gamma_j}. \]
Recall \( \omega_{ij} \) captures the expenditure share of sector \( i \) on good \( j \), or the *equilibrium importance* of \( j \) as a supplier for \( i \). On the other hand, \( \hat{\omega}_{ij} \) captures the share of good \( j \) that is used by sector \( i \) as a fraction of total output of sector \( j \). In other words, \( \hat{\omega}_{ij} \) captures the *equilibrium unconstrained* (or technological) *importance* of \( j \) as a supplier for \( i \) and \( \hat{\omega}_{ij} \) as a consumer for \( j \). To interpret these measures in another way, recall that an input-output table contains pair-wise flow of value between industries. The input-output coefficient matrix \( \Omega \) is obtained by dividing IO table entries by the output of using industry, whereas \( \hat{\Omega} \) is obtained by dividing entries by the output of supplying industry, and similarly for \( \Sigma \) and \( \hat{\Sigma} \).
Proposition 1.6. In equilibrium,

\[ \xi' \propto 1' \left( I - \Omega \right) \left( I - \Sigma \right)^{-1}. \]

Hirschman (1958) argues that industrial policies should target and promote the economic sectors with the strongest linkages. Following his work, there is a literature that aims to develop measures of Hirschmanian linkages and use these measures to study economic policies, including Chenery and Watanabe (1958), Rasmussen (1965), Yotopoulos and Nugent (1973), and many others. All measures proposed and debated during that period are ad-hoc and without microfoundations. One notable measure by Jones (1976) is later used for applied policy work:

\[ \delta^\text{Jones}_i = \left( 1' \left( I - \Sigma \right)^{-1} \right)_i. \]

\( \delta^\text{Jones}_i \) is proposed as a measure of the “forward linkages” for industry \( i \) and is supposed to capture the extent to which downstream industries could benefit from output gains in industry \( i \). Noland (2004) finds that this measure strongly explains which sectors were promoted during South Korea’s “Heavy-Chemical Industry” drive in the 1970s. It turns out that \( \delta^\text{Jones}_i \) is closely related to sales gap under a particular assumption of how credit constraints affect sectoral production. Specifically, suppose credit constraints in the economy create a constant wedge \( \varphi \) between output elasticities and expenditure shares for all industries and all intermediate inputs such that \( \omega_{ij} = \sigma_{ij} \varphi \) for all \( i, j \). This assumption corresponds to a setup in which all intermediate inputs are constrained in every sector, and the allocation of working capital equalizes the sectoral average private marginal returns to capital inputs. In this scenario,

\[ \xi \propto \text{const} + \delta^\text{Jones}. \]

That is, under the assumption that \( \omega_{ij} = \sigma_{ij} \varphi \) for all \( i \) and \( j \), the sales gap measure is an affine transformation of \( \delta^\text{Jones} \).

Empirics

I now turn to examining whether the sufficient statistic I have developed can explain interventions implemented by developing countries. The two essential ingredients in constructing the sales gap measure are sectoral sales and influence. Sectoral sales is directly observable from national accounts, but influence requires estimation. If TFP shocks are directly observable, sectoral influence can be estimated from a regression of aggregate output on the TFP shocks. On the other hand, TFP shocks are almost never observed and are often obtained as residuals from estimations of aggregate production functions. For this reason, I choose an indirect route to estimating influence by first estimating the matrix of production elasticities, i.e. the unconstrained input-output table, and
then computing the influence vector using the Leontief-inverse formulae provided in early sections. I perform two distinct empirical exercises that recover the sales gap measure via this route, and I examine the correlation between a sector's sales gap and plausible proxy measures of government interventions therein.

1.3 Structural Analysis: China

In this section I conduct empirical analysis in the context of China. Because it has a government with deep fiscal capacity and a strong legacy of state intervention in production due to its socialist roots, China is a particularly interesting setting in which to apply my theoretical results. I estimate production elasticities using firm-level production data from year 2007 of China’s Annual Industrial Survey (AIS) for 66 manufacturing industries and construct the sales gap measure by applying these estimates to the 2007 Chinese input-output table. I show that private firms in sectors with higher sales gaps tend to receive more external loans and pay lower interest rates, and that the sectoral presence of SOEs is heavily directed towards sectors with larger sales gaps rather than sectors that are the most constrained or have the highest private return to capital.

History of SOEs in China

SOEs play two important roles in this part of my empirical analysis. First, I treat the presence of SOEs as a measure of government intervention into a sector, as they can be an indirect vehicle for the state to subsidize production when direct production subsidies are difficult to implement due to practical obstacles. Second, I exploit SOEs in the estimation strategy by assuming that, conditional on being in a sector, they are price-taking profit maximizers that are unconstrained but potentially less productive than their private-owned counterparts, as in Song et al. (2011). In what follows, I provide a very brief overview of Chinese SOE’s institutional setting to defend this assumption.

The Chinese manufacturing sector was dominated by SOEs before the late 1990s. Under a “dual track” system, private and state-owned firms coexisted but were subject to vastly different market regulations. SOEs faced price controls and production quotas, and most importantly, profit maximization was not their objective. Managers were typically ranking Communist party officials whose compensation was heavily regulated. Moreover, the government’s top priorities include promoting social stability and avoiding layoffs, and loss-making SOEs were kept alive by loan injections and bank bailouts. The lack of exit and market selection of more productive firms further distorted SOEs’ economic incentives.

4 Indeed this is a popular view among economic historians on Taiwan’s industrial policies in 1960s, where public enterprises were heavily involved in the manufacturing sector (Hernandez 2004).

5 Detailed discussions of the history of SOEs and their role in the Chinese economy can be found in Naughton (2006), Brandt et al. (2008), Zhu (2012), Storesletten and Zilibotti (2014), and Hsieh and Song (2015).
Major waves of SOE reform began in the late 1990s, with the explicit objective of letting SOEs compete with private firms through market mechanisms while keeping state control. The Chinese government retracted its commitment to stable employment, and a large number of failing SOEs were closed down or privatized through management buyouts. While the overall SOE share in the manufacturing sector steadily declined between 1998 and 2007, surviving SOEs are positively selected on both size and productivity. As a result of the reforms, many of the larger SOEs became corporatized and are now publicly traded (though the state retains controlling shares), successful market participants. In fact, many scholars of the Chinese economy argue that the SOE reforms were intended to turn SOEs into profit-maximizing entities (Hsieh and Song 2015). Because my empirical setting focuses on the industrial production of year 2007, well into the post-reform period, it may therefore be reasonable to model SOEs as profit maximizers.

There is also well-documented evidence that SOEs in China receive easy access to credit from state-owned Chinese banks (Boyreau-Debray and Wei 2005), while private firms face significant financial constraints (Allen et al. 2005), especially for capital goods purchases (Dollar and Wei 2007, Riedel et al. 2007). Song et al. (2011, section I.C) provide a systematic compilation of related evidence.

**Technological Assumptions** A significant data challenge is that there is no firm-level data on input use by sector of origin. Instead, the available firm-level production surveys (AIS) categorize inputs into labor, capital, and intermediate materials, as do most similar manufacturing data sets for other countries. To make progress despite the data limitations, I partition industries into those that produce “capital goods” and those that produce “intermediate materials.” I further assume that firm production functions are separable in these two goods bundles and are homogeneous within each bundle. This assumption ensures that expenditures on each bundle of intermediate goods are comparable across firms within the same industry, as the ratio of expenditure between any two goods within a bundle is the same across firms within the same industry. Formally,

**Assumption 1.2.** Intermediate goods can be partitioned into mutually exclusive categories $K$ and $X$ such that production functions in all industries are homothetically separable in the two groups of industries. That is, for all industries $i = 1, \cdots S$, there exist continuously differentiable functions $k_i : \mathbb{R}^{|K|}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ and $x_i : \mathbb{R}^{|X|}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ that are homogeneous of degree one, such that

$$f_i(\ell, m_1, \cdots, m_S) = f_i(\ell, k_i(\{m_j\}_{j \in K}), x_i(\{m_j\}_{j \in X})).$$

Each industry is assigned to be a producer of capital goods iff more than 5% of its total output in 2007 is used for “gross capital formation”, which is a measure that captures the value of goods that is un-depreciated in the accounting year and is to be used at a future time. Out of 66 industries, 20 are assigned as capital good producers, and 46 are assigned as material good producers. Our results are qualitatively insensitive to alternative cutoff rules.
Under the homogeneity assumptions on \( k_i (\cdot) \) and \( x_i (\cdot) \), there exist industry-specific price indices \( p_i^K \) and \( p_i^X \) for the bundle of capital and material goods, respectively:

\[
p_i^K = \min_{\{m_j\}_{j \in K}} \sum_{j \in K} p_j m_j \quad \text{s.t.} \quad k_i \left( \{m_j\}_{j \in K} \right) \geq 1
\]

\[
p_i^X = \min_{\{m_j\}_{j \in X}} \sum_{j \in X} p_j m_j \quad \text{s.t.} \quad x_i \left( \{m_j\}_{j \in X} \right) \geq 1
\]

The unconstrained profit maximization problem of SOEs can be re-written as

\[
\max_{\ell, k, x} p_i z_i (\nu) f_i (\ell, k, x) - w \ell - p_i^K k - p_i^X x,
\]

while private firms solve the same problem but are subject to the credit constraints. Motivated by the evidence that private firms in China obtain significantly fewer bank loans when financing capital investments (Dollar and Wei 2007, Riedel et al. 2007) and have substantially lower capital-output and capital-labor ratios than SOEs (Song et al. 2011), I specify that only capital goods are subject to credit constraints, while labor and intermediate materials are not:

\[
p_i^K k \leq W_i (\nu). \quad (1.31)
\]

The fact that capital inputs are subject to the constraint implies, as shown in the previous section, that there will be a firm-level wedge between the expenditure share on capital and output elasticity for private firms. In Appendix 1.11.2 I conduct a specification test for this assumption\(^6\).

To rationalize the residuals in estimation, I introduce an ex-post and multiplicative productivity shock \( \varepsilon_i (\nu) \) that affects output but is observable to firms only after input choices are made. Formally, the productivity \( z_i (\nu) \) that affects input choices can be written as the product between a component that is known to the firm when making input choices, \( \bar{z}_i (\nu) \), and the expectation over the ex-post productivity shock:

\[
z_i (\nu) = \bar{z}_i (\nu) \mathbb{E}_\nu [\exp (\varepsilon_i (\nu))]
\]

The observable firm revenue \( r_i (\nu) \) can be written as

\[
p_i q_i (\nu) = p_i z_i (\nu) \frac{\exp (\varepsilon_i (\nu))}{\mathbb{E}_\nu [\exp (\varepsilon_i (\nu))]} f_i (\ell_i (\nu), k_i (\nu), m_i (\nu)). \quad (1.32)
\]

The term \( \frac{\exp (\varepsilon_i (\nu))}{\mathbb{E}_\nu [\exp (\varepsilon_i (\nu))]} \) can be equivalently interpreted as a multiplicative measurement error in the firm's revenue. In what follows, I choose a normalization of \( \varepsilon_i (\nu) \) such that \( \mathbb{E}_\nu [\exp (\varepsilon_i (\nu))] = 1. \)

\(^6\)Specifically, I show that the estimated wedges between firm-level expenditure shares and output elasticity are much higher for capital inputs than for labor and material inputs, and the average wedge across industries on labor and material inputs is close to one, which is the unconstrained benchmark.
assume the distribution of the ex-post shocks (or measurement errors) is independent of both the firm’s status and any other variable in the firm’s information set when making input choices to ensure that the ex-post shocks are not correlated with input choices.

**Assumption 1.3.** $\epsilon_i(\nu)$ is independent of $z_i(\nu), W_i(\nu)$, and the firm’s status (SOE or private).

### 1.3.1 Identification

**Identification of Output Elasticity with Respect to Capital**  I now lay out the identification strategy for firm-level output elasticity with respect to the bundle of constrained inputs, or capital $K$. For this subsection, I temporarily drop industry subscript $i$ for the ease of notation and focus the exposition on a single industry. The same argument applies to all industries.

The main difficulty in estimating features of production functions comes from the transmission bias: the Hicks-neutral productivity shock $z(\nu)$ endogenously affects both firm-level output and the observable inputs choices, and therefore a regression of firm-level output on inputs yields biased estimates of output elasticities. To get around this problem, I adapt the strategy proposed by Gandhi et al. (2016) to my static setting. The key insight from Gandhi et al. (2016) is that for unconstrained firms, the first-order conditions with respect to inputs provide additional information that enables us to purge the endogeneity generated by Hicks-neutral productivity $z(\nu)$. More specifically, under the assumption that SOEs are unconstrained, an SOE’s first-order condition with respect to capital is

$$ p^K = pz(\nu) \frac{\partial f(\ell(\nu), k(\nu), x(\nu))}{\partial k}. \quad (1.33) $$

Multiplying both sides of (1.33) by $k(\nu)$, dividing by revenue using (1.32), and then taking logs,

$$ \ln s^K(\nu) = \ln \sigma^K(\ell(\nu), k(\nu), x(\nu)) - \epsilon(\nu), \quad (1.34) $$

where $s^K(\nu) = \frac{p^K_k(\nu)}{r(\nu)}$ is the capital share and $\sigma^K(\nu) \equiv \frac{\partial}{\partial \ln k} \ln f(\ell(\nu), k(\nu), x(\nu))$ is the output elasticity with respect to capital goods. Given that $s^K(\nu), w\ell(\nu), p^K_k(\nu)$, and $p^X x(\nu)$ are all observable and that the ex-post productivity shock $\epsilon(\nu)$ is independent of the input expenditures,

**Proposition 1.7.** The output elasticity with respect to the bundle of capital goods $\sigma^K(\cdot)$ is non-parametrically identified from SOE production data. The distribution function $F^*(\cdot)$ of the ex-post productivity shock $\epsilon(\nu)$ is also non-parametrically identified.

**Identification of Return to Capital for Private Firms** Next, I take $\sigma^K(\cdot)$ as a known function of input expenditures and outline how this information can be used to identify the distribution of firm-level returns to capital for the private firms. Taking the first-order condition with respect to capital inputs and multiplying it by $k(\nu)/pq(\nu)$ again derives an equation relating the private...
firm's expenditure share on capital goods to output elasticities:

\[ s^K(\nu) = \sigma^K(\nu) \varphi^K(\nu) e^{-\varepsilon(\nu)}, \]

The left-hand side is again directly observable, and \( \sigma^K(\nu) \) is a known function of observable variables thanks to the previous identification result. Therefore for any given firm \( \nu \),

\[ \left( \varphi^K(\nu) \right)^{-1} e^{\varepsilon(\nu)} = \frac{\sigma^K(\nu)}{s^K(\nu)} \]

and the distribution function of \( \left( \varphi^K(\nu) \right)^{-1} e(\nu) \) is identified.

While I cannot separately recover firm-level wedges \( \left( \varphi^K(\nu) \right)^{-1} \) and the ex-post productivity shock \( e(\nu) \) for each individual firm \( \nu \), I can indeed recover distributional properties of \( \varphi^M(\nu) \). Specifically, because the distribution function of both \( \ln \varphi^M(\nu) - e(\nu) \) and \( e(\nu) \) are identified, I can apply the method of deconvolution (Chen et al. 2011) and recover the distribution function of \( \varphi^M(\nu) \) such that:

**Proposition 1.8.** The distribution function \( F^{\varphi^M}() \) of \( \varphi^M \) is non-parametrically identified.

### 1.3.2 Estimation

In this subsection I outline an estimation procedure to construct the sales gap measure \( \xi \) based on the identification results from the previous subsection.

From the AIS micro data, I observe firm revenue, costs of labor inputs, cost of intermediate materials, and the book value of total firm-level capital stock. Firm production in the real world is dynamic in nature, and one can interpret the one-shot production game after entry in the static model as representing the steady-state of a dynamic production game. In order to map the dynamic real-world data to the static production model, one should adopt a “flow” capital expenditure measure that corresponds to the gross capital depreciation during the production period. Reliable measures of firm-level capital depreciation are unavailable in the Chinese production data, but fortunately, identifying output elasticities with respect to capital only requires the econometrician to observe a constant multiple of the capital expenditure, which can be proxied by the book value of capital stock variable as long as I assume firms within an industry have the same depreciation rate.

Formally, for each firm \( \nu \) in industry \( i \), I separately observe firm-level revenue \( r_i(\nu) = p_i q_i(\nu) \) and expenditures on labor \( w \ell_i(\nu) \), on the bundle of commodity goods \( p^X_i x_i(\nu) \), and a constant multiple of the expenditures on capital goods \( c_i \cdot p^K_i k_i(\nu) \), where the industry-specific constant \( c_i \) is the inverse of the depreciation rate. The identification results in the previous subsection specify
\(\sigma^K(\nu)\) as a function of input quantities. On the other hand, given that the depreciation rate and input prices are industry wide, the observed variables are simply multiplicative transformations of input quantities, and \(\sigma^K(\nu)\) can be re-written as a function of the observed variables via a simple change of variable\(^7\). To avoid carrying redundant notations, I simply write \(\ell_i(\nu), x_i(\nu), k_i(\nu)\) as observed variables and omit the multiplicative constants.

For each industry \(i\), I parametrize \(\sigma^K_i(\cdot)\) as a second-order polynomial in logs of the input expenditures:

\[
\ln s^K_i(\nu) = \log \left( \eta^1_i + \eta^2_i \cdot \log \ell_i(\nu) + \eta^3_i \cdot \log k_i(\nu) + \eta^4_i \cdot \log x_i(\nu) \right. \\
+ \eta^5_i \cdot (\log \ell_i(\nu))^2 + \eta^6_i \cdot (\log k_i(\nu))^2 + \left. \eta^7_i \cdot (\log x_i(\nu))^2 \right) - \epsilon_i(\nu) \tag{1.35}
\]

I estimate the model on the sample of SOEs for each industry \(i\)\(^8\), using GMM with moment conditions \(E, [\exp(\epsilon_i(\nu))] = 1\) (our chosen normalization of \(\epsilon_i(\nu)\)) and \(Cov(\epsilon_i(\nu), v_i(\nu)) = 0\) for \(v \in \{\log \ell, \log k, \log x, (\log \ell)^2, (\log k)^2, (\log x)^2\}\). In Appendix 1.11.2 I show that my results are qualitatively similar if I simply parametrize \(\sigma^K_i(\cdot)\) as an industry-specific constant, which corresponds to production functions with constant output elasticity of capital inputs.

Equipped with estimates \(\hat{\sigma}^K_i(\nu)\), I obtain the residuals for private firms:

\[
\widehat{\text{res}}_i(\nu) = \frac{\hat{\sigma}^K_i(\nu; \hat{\eta}_i)}{\hat{\sigma}^K_i(\nu)} = \hat{\varphi}^K_i(\nu)^{-1} \hat{\epsilon}_i(\nu). \tag{1.36}
\]

The average (weighted) private return to capital inputs in industry \(i\) can be written as

\[
\left( \hat{\varphi}^K_i \right)^{-1} = \frac{E_{\nu} \left[ \varphi^K_i(\nu)^{-1} k_i(\nu) \right]}{E_{\nu} \left[ k_i(\nu) \right]} = \frac{E_{\nu} \left[ \varphi^K_i(\nu)^{-1} \epsilon_i(\nu) k_i(\nu) \right]}{E_{\nu} \left[ \epsilon(\nu) \right] E_{\nu} \left[ k_i(\nu) \right]}, \tag{1.37}
\]

where the second line follows by independence in Assumption 1.3.

I estimate \(\left( \hat{\varphi}^K_i \right)^{-1}\) by forming estimates of the three expectations separately using an empirical Bayes procedure (Morris 1983) that exploits cross-industry information. In particular, \(E_{\nu} \left[ k_i(\nu) \right]\)

\(^7\)That is, I can define a revenue production function by simply scaling the quantity variables by constants \((p, w, c^K, p^K)\):

\[
f_R \left( w \ell(\nu), c^K k(\nu), p^K x(\nu) \right) \equiv pf(\ell(\nu), k(\nu), x(\nu)) \text{ with } \frac{\partial \ln f_R(\nu)}{\partial \ln (c^K k(\nu))} = \frac{\partial \ln f(\nu)}{\partial \ln (k(\nu))}.
\]

\(^8\)Appendix Table 1.11.7 provides summary statistics on the number of SOEs in each industry.
and $\mathbb{E}_{\nu}\left[\varphi^K_i(\nu)^{-1}e_i(\nu)k_i(\nu)\right]$ are estimated from private firms' capital stock $k_i(\nu)$ and estimation residuals $\hat{r}_i(\nu)$, whereas $\mathbb{E}_{\nu}[\hat{\epsilon}(\nu)]$ is estimated using the residuals for the SOEs. I discuss the procedure in detail in Appendix 1.10. I also estimate an unweighted average private return to capital for each industry, $\mathbb{E}_{\nu}\left[(\varphi^K_i(\nu))^{-1}\right]$, via a similar procedure.

Next, I use the estimates $(\hat{\varphi^K_i})^{-1}$ to construct the input-output elasticity matrix. Because only capital goods are subject to credit constraints, those entries in the observed input-output matrix, which represent expenditure shares on intermediate materials, do reflect actual elasticities, while to obtain elasticity with respect to capital goods I multiply the observed expenditure shares with the input-using industry's average private return:

$$\Sigma_{ij} = \begin{cases} (\varphi^K_i)^{-1}\Omega_{ij} & \text{if } j \in K \\ \Omega_{ij} & \text{if } j \in X, \end{cases}$$ (1.38)

where recall $\Sigma$ denotes the elasticity matrix and $\Omega$ denotes the matrix of expenditure shares or the observed input-output table. Lastly, I construct the industry level sales gap measure $\hat{\xi}_i$ by

$$\hat{\xi}_i = \hat{\mu}_i/\gamma_i,$$ (1.39)

where

$$\hat{\mu}' = \frac{\beta' (I - \hat{\Sigma})^{-1}}{\beta' (I - \Sigma)^{-1} \alpha}, \quad \gamma' = \frac{\beta' (I - \Omega)^{-1} \alpha}{\beta' (I - \Omega)^{-1} \alpha},$$ (1.40)

and as guided by my theory (Corollary 1.2), industry level social return to expenditure on capital is defined as

$$\overline{SR}_i^K = \hat{\xi}_i \times (\varphi^K_i)^{-1}.$$ (1.41)

### 1.3.3 Data

The firm-level variables come from year 2007 of China's Annual Industrial Survey (AIS), which is a firm-level manufacturing survey that includes private firms with sales above 5 million RMB as well as all SOEs. To make the two sets of firms comparable, I drop SOEs with sales below the 5 million RMB cutoff. The 2007 Chinese input-output table comes from the national accounts published by China's National Bureau of Statistics. Following the methodology described by Hsieh and Song (2015), throughout this section I use the term SOE to refer both to firms that are legally registered as state-owned and to firms whose controlling shareholder is the state. Below, I elaborate on three details of how the conceptual framework is mapped to actual data and estimation.

The first issue pertains to the construction of the sales vector $\gamma$. Rather than using actual sales
directly observed from national accounts for my analysis, I use the measure constructed from the Leontief inverse of the input-output table, with the final share vector $\beta$ measured as

$$
\beta_i = \frac{\text{Private and public consumption of good } i}{\text{Total priv. and public consumption of all goods}}.
$$

(1.42)

In a closed and static economy, sectoral sales directly observed from national accounts are mechanically proportional, by construction, to the measure I adopt. On the other hand, real-world production involves both dynamic accumulation of capital and inventory as well as imports and exports, hence the observed total sectoral sales are equivalent to that constructed based on the Leontief inverse method if the final shares are modified to incorporate net sectoral output that is used for dynamic accumulation and trade. I exclude these components of sectoral output from the final share measure in equation (1.42) because they are not relevant to my theoretical model, and I use the same final share vector to construct both influence and sales vector according to (1.40). I show in Appendix Table 1.11.1 that my results are robust to using influence and sales measures constructed with alternative final share vectors that include net exports.

The second issue relates to constructing net value-added share $\alpha$ and how I deal with accounting profits and the total consumption of fixed capital accumulated in previous accounting years. Together with wage payments, these entries are recorded in input-output tables as being part of each industry’s gross value-added. In my model, there are accounting profits but no economic profits: the accounting profits cover the fixed cost of entry. For this reason, I apply the same treatment to profits in the model and take the net value-added of an industry as being the recorded wage,

$$
\alpha_i = \frac{\text{Wage}_i}{\text{Output}_i}.
$$

(1.43)

The recorded value for the total consumption of previously accumulated fixed capital is proportionally added to the value of capital inputs used for production. These recorded entries for capital depreciation constitute a small fraction of gross value-added, averaging 12% for the manufacturing sectors and 13% for the economy as a whole, and account for an even smaller fraction of total output (2.5% on average for the manufacturing sector). As a result, adjusting for the value of capital goods used for production has little impact on the results.

The third issue pertains to how I combine the micro-estimates from the AIS data with the Chinese input-output table when constructing the sales gap measure. The input-output table records the flow of value across manufacturing as well as primary and tertiary sectors, while the AIS micro data cover only manufacturing industries under a different and more disaggregated industrial classification code. I manually create the concordance between manufacturing industries in the two data sources, merging industries when necessary, with a final combined data set that includes 66 distinct manufacturing industries for which wedges can be recovered from the AIS data. My empirical strategy does not recover the wedges for the primary and tertiary sectors due to the lack of micro data, and I report results in the main text based on the conservative assumption
that firms in these sectors are not constrained. In Appendix Table 1.11.2 I show that my results are robust to dropping the primary and tertiary sectors altogether and performing the Leontief inverse on a partial input-output table with the manufacturing sectors only. I also winsorize the estimates of \((\hat{\varphi}_i^K)^{-1}\) below one to reflect the theoretical lower bound, as \((\hat{\varphi}_i^K)^{-1}\) should be equal to one if firms in industry \(i\) are unconstrained and greater than one if a positive measure of firms are constrained. This winsorization affects only 7 out of the 66 manufacturing industries, thus indicating that private firms in most industries face constraints in their capital choices, and the fact that these 7 industries have wedge estimates less than one could be due to sampling error.

1.3.4 Results

Private and Social Return to Capital Table 1.1 provides summary statistics for the estimated wedges, social return, and sales gap measures. Column (1) corresponds to estimates of the unweighted sectoral average of firm-level wedges, \(E \left[ \varphi_i^K (\nu)^{-1} \right]\). Across sectors, an average firm has gross private rate of return to capital of 1.2 or net rate of return of 20%, which is inline with other estimates from the literature (e.g. Bai et al. 2014). Column (2) corresponds to estimates of \((\hat{\varphi}_i^K)^{-1}\), the sectoral average private return to capital inputs weighted by the amount of capital used by each firm. This weighted average return is the relevant object in our model (in equation 1.18 and 1.38) and can be written as the sum between the unweighted average return \(E \left[ \varphi_i^K (\nu)^{-1} \right]\) and the covariance between the level of and return to capital inputs:

\[
(\hat{\varphi}_i^K)^{-1} = E_i \left[ (\varphi_i^K (\nu))^{-1} \frac{k_i (\nu)}{E_i (k_i (\nu))} \right] = E_i \left[ (\varphi_i^K (\nu))^{-1} \right] + Cov_i \left[ (\varphi_i^K (\nu))^{-1}, \frac{k_i (\nu)}{E_i (k_i (\nu))} \right].
\]

The estimates \((\hat{\varphi}_i^K)^{-1}\) averages to 1.34 across sectors and is higher than the unweighted average \(E_i \left[ (\varphi_i^K (\nu))^{-1} \right]\) in every industry, thereby implying that the covariance term is positive and that firms using more capital inputs also tend to have higher marginal return to additional capital inputs. This empirical finding implies that while more productive firms are allocated more credit \((z_i (\nu)\) and \(W_i (\nu)\) are positively correlated), these firms should receive even more working capital in order for marginal return of capital inputs to equalize across firms within each sector. This finding is consistent with the evidence in Hsieh and Klenow (2009) that larger firms in China tend to have higher returns to capital.

The network effect of the sectoral credit constraints is large: while the \((\hat{\varphi}_i^K)^{-1}\) averages to 1.34, or 34% net private return to capital, the network work inefficiency magnifies the average gross social return to capital inputs to 1.46. Much of this large difference of 12 percentage points is due to a heavier right tail induced by the network adjustment in constructing the social return (equation 1.41), as illustrated by Figure 1-2, which shows the density distributions of \((\hat{\varphi}_i^K)^{-1}\) and \(\bar{S}R_i^K\) with
dashed-grey and solid-black lines, respectively.

Table 1.1: Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>$E \left[ \hat{\varphi}_i^K (\nu)^{-1} \right]$</th>
<th>$(\hat{\varphi}_i^K)^{-1}$</th>
<th>$\hat{\xi}_i$</th>
<th>$\hat{SR}_i^K$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1.20</td>
<td>1.34</td>
<td>1.08</td>
<td>1.46</td>
</tr>
<tr>
<td>St. dev</td>
<td>0.19</td>
<td>0.24</td>
<td>0.17</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Notes: the four columns respectively show summary statistics for
the following industry-level variables: 1) unweighted average of
firm-level marginal returns to capital inputs in the industry; 2)
"private return to capital inputs", i.e. average firm-level marginal
returns to capital inputs, weighted by the capital stock of each
firm; 3) the sectoral sales gap; 4) the social return to capital
inputs.

Figure 1-2: Density of sectoral private return $(\hat{\varphi}_i^K)^{-1}$ and social return $\hat{SR}_i^K$

**Sectoral Sales Gap** Table 1.2 lists the top and bottom 15 manufacturing industries ranked by
their sales gap measures. Unsurprisingly, many of the industries characterized as capital goods
producers, such as “industrial furnace and boiler” and “rail equipment”, appear on the Top 15 list,
i.e. top sales gaps. This is because these industries produce goods that are directly subject to credit
constraints, which distort their sales below optimal production level. What is perhaps less obvious
is the set of metal industries that are also on this Top 15 list, such as the smelting, stamping, and
rolling of iron, steel, and ferrous and non-ferrous alloy. Because these industries are not classified as
capital goods by the selection criterion, the purchase of these goods are not directly subject to credit constraints. The reason that these goods are on the Top 15 list is because of network propagation of financial distortions: the metal products are extensively used as intermediate materials that go into the production of capital goods, mostly machines and equipment. Credit constraints distort the sales of capital goods downwards, which in turn distorts the equilibrium sales of the metal products through upstream propagation of the demand distortion. As a result, manufacturers of these unconstrained inputs end up with high sales gaps, and the economy could benefit from subsidizing their production.

On the other side of the table, light industries—those that produce food and textiles—dominate the Bottom 15 list. These industries are downstream from the capital goods and are arguably also the most downstream in the network structure: although they use capital goods for production, these industries’ output is used more for direct consumption rather than industrial production, and, in particular, their output is not heavily used for the production of capital goods. While private firms in these industries do face credit constraints that lower their effective productivity, the constraints do not impose large sales gaps, and, in aggregate, these are the sectors that are too large relative to optimum. Lastly, the set of industries that are not on these partial lists includes chemical industries and a set of industries that produce non-metallic materials such as rubber, lime glass, and plastic products. Despite the fact that “upstreamness” is not uniquely defined in the context of a complete network in which every industry purchases at least some inputs from every other, we tend to think of these as the midstream industries: these products are used more as intermediate inputs to manufacturing rather than for final consumption, but, on the other hand, the production of these goods requires more upstream machineries than the extent to which these goods are used as inputs for upstream production.

Debt-To-Capital Ratio and Interest Rates  I next examine the correlation between sales gap and plausible measures of government intervention. Though the manufacturing sector in China was largely driven by market forces in 2007, the credit market remained predominantly state-controlled, with the government holding direct ownership of the largest commercial banks. Targeted and subsidized credit through the banking sector to both SOEs and private manufacturing firms played an important policy role (Aghion et al. 2015). While detailed firm-level data on bank loans is unavailable, firms in the AIS do report their total interest payments and total liabilities. Based on these variables, I derive proxy measures of credit market interventions. I define the Debt-To-Capital ratio \((D/K)\) for firm \(\nu\) in industry \(i\) as

\[
(D/K)_{i}(\nu) \equiv \frac{\text{Total Liabilities}_{i}(\nu)}{\text{Total Capital Inputs}_{i}(\nu)}
\]
Table 1.2: Top and Bottom 15 Industries Ranked by Sales Gap

<table>
<thead>
<tr>
<th>Top 15</th>
<th>Bottom 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industrial furnace and boiler</td>
<td>Misc. food products</td>
</tr>
<tr>
<td>Metal cutting machinery</td>
<td>Meat processing</td>
</tr>
<tr>
<td>Misc. general-purpose machinery</td>
<td>Medicine manufacturing</td>
</tr>
<tr>
<td>Auto manufacturing</td>
<td>Sugar making</td>
</tr>
<tr>
<td>Electrical machinery</td>
<td>Liquor and alcoholic drinks</td>
</tr>
<tr>
<td>Fabricated metal products</td>
<td>Clothing and footwear manufacturing</td>
</tr>
<tr>
<td>Ferroalloy smelting</td>
<td>Stationery manufacturing</td>
</tr>
<tr>
<td>Iron smelting</td>
<td>Household chemical products</td>
</tr>
<tr>
<td>Misc. special-purpose equipment</td>
<td>Knit textiles</td>
</tr>
<tr>
<td>Steel smelting</td>
<td>Consumer electronics</td>
</tr>
<tr>
<td>Misc. electrical equipment</td>
<td>Leather, fur, and down products</td>
</tr>
<tr>
<td>Steel rolling and stamping</td>
<td>Vegetable oil manufacturing</td>
</tr>
<tr>
<td>Non-ferrous metal rolling and</td>
<td>Seafood processing</td>
</tr>
<tr>
<td>stamping</td>
<td>Wool weaving and printing</td>
</tr>
<tr>
<td>Rail equipment</td>
<td>Cotton and polyester weaving and printing</td>
</tr>
<tr>
<td>Non-ferrous metal smelting</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table ranks 66 manufacturing sectors by their sectoral sales gaps and shows the top and bottom 15 sectors.

and I define firm-level interest rate as

\[ \text{IntRate}_i(\nu) = 100 \times \frac{\text{Total Interest Payment}_i(\nu)}{\text{Total Liabilities}_i(\nu)}. \]  

(1.44)

Holding private marginal return to capital \( (\nu_i^K)^{-1} \) constant, the social return to capital inputs is higher in sectors with higher sales gaps. Welfare-enhancing policies should direct more working capital towards these sectors, potentially through a combination of targeted and subsidized lending. While sectors that have higher average \( D/K \) ratios across firms can be interpreted through the model as those that tend to receive more external loans\(^9\), interest rates do not have a natural counterpart in the model because credit rationing is common in lending markets in China and interest rates do not clear the credit market. Nevertheless, lower average sectoral interest rates can be viewed as evidence of subsidized credit to the sector.

Table 1.3 compares the mean of these two measures of credit market intervention between private firms and SOEs. Columns (1) and (3) respectively regress the firm-level \( D/K \) ratio and interest rates on a dummy that captures whether the firm is state-owned, and columns (2) and (4) control for sector fixed effects for the respective outcome variable. The results reveal that SOEs receive significantly more favorable access to credit markets and are on average 9.1 percentage points higher.

\(^9\) The total working capital \( W_i(\nu) \) can be modeled as the sum of entrepreneurial wealth \( EW_i(\nu) \) and external liability \( D_i(\nu) \) and we have \( (D/K)_i(\nu) = \frac{D_i(\nu)}{W_i(\nu)} \). When \( EW_i(\nu) \) do not vary systematically across sectors, firms in sectors with higher average \( D/K \) ratio tend to receive more external loans.
in debt-to-capital ratio, compared to a baseline of 54 percentage points for the private firms. SOEs also pay interest rates that are 2.1 percentage points lower relative to a baseline of 4.3 percentage points paid by their private counterparts\(^{10}\). These results are consistent with the assumption that SOEs are not subject to credit constraints when making production choices.

Tables 1.4 and 1.5 explore the correlation between sales gap and the two measures of credit market interventions using the following firm-level regression:

\[
\text{Outcome}_{i} (v) = \delta^1 + \delta^2 \times \xi_i + \delta^3 \times (\bar{s}_{i}^{K})^{-1} + \text{Controls}_i + \epsilon_i (v).
\]

Table 1.4 shows the outcome variable Debt-To-Capital ratio and Table 1.5 shows the firm-level interest rate. Columns (1) through (4) in both tables show results estimated on the sample of private firms, whereas the results in columns (5) through (8) are based on the sample of SOEs. All standard errors are clustered conservatively at the industry level.

A consistent pattern that emerges from Tables 1.4 and 1.5 is that private firms in sectors with higher sales gaps tend to have higher \(D/K\) ratios and tend to pay lower interest rates, suggesting that private firms in these sectors receive favorable access to credit markets. Moreover, neither of these intervention measures correlates with the average sectoral wedge \((\bar{s}_{i}^{K})^{-1}\), suggesting that it is the network inefficiency measure (sales gap) rather than the private marginal return to capital that predicts the cross-sector variation in credit market interventions.

The preferred specification in column (4) of both tables includes controls for firm-level capital intensities as well as the fraction of sectoral output that is used to form capital stock for future production. The former variable \((K/Y)\) measures how intensively capital is used for production, while the latter \((\text{CapForm})\) measures the capital content of sectoral output and is the basis on which sectors are categorized as producing capital or material goods (c.f. Assumption 1.2). These controls are introduced to partially address the concern that \(D/K\) ratio might be mechanically driven by sectoral characteristics. While firms with higher capital intensities tend to have better access to external loans and receive lower interest rates, these control variables do not remove the correlation between sales gap and credit intervention measures: private firms in sectors with one percentage point higher sales gaps tend to have 0.0924 percentage point higher \(D/K\) ratio and tend to receive interest rates that are 0.0247 percentage point lower. Given that the standard deviation in the average sectoral \(D/K\) ratio and interest rate are 4.67 and 1.53 percentage points, respectively, these effects are quantitatively large: when comparing across sectors, those that are one standard deviation higher in sales gaps (24 percentage point as in Table 1.1) tend to have average \(D/K\) ratios that are 0.47 standard deviations higher and average interest rates that are

\(^{10}\)For interpretational purposes, I treat firms that report zero interest payments as missing data in the main text; all results that involve the interest rate variable are based on the subsample of firms that report positive total interest payments, which corresponds to 58% of firms in the full sample. In Appendix 1.11.1, I show that my results are robust to using the entire sample of firms and interpreting firms with no recorded interest payment as those facing zero interest rate.
Table 1.3: SOEs have higher Debt-To-Capital ratios and pay lower interest rates

<table>
<thead>
<tr>
<th></th>
<th>$(D/K)_i (\nu)$</th>
<th>$IntRate_i (\nu)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$(SOE_i (\nu))$</td>
<td>0.0912***</td>
<td>0.0937***</td>
</tr>
<tr>
<td></td>
<td>(0.00227)</td>
<td>(0.00228)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.544***</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.000505)</td>
<td>-</td>
</tr>
<tr>
<td>Sector Fixed Effects</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Obs.</td>
<td>296784</td>
<td>296784</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.005</td>
<td>0.027</td>
</tr>
</tbody>
</table>

Notes: The table compares the firm-level Debt-to-Capital ratio (columns 1 and 2) and interest rate (columns 3 and 4) of private firms and SOEs. Columns (2) and (4) add sector fixed effects to columns (1) and (3), respectively. Columns (1) and (2) drop outlier firms with Debt-To-Capital ratios that are either negative or above the 99th percentile, and columns (3) and (4) drop firms with interest rate that is either negative or above the 99th percentile.

0.39 standard deviation lower.

Columns (5) through (8) in Tables 1.4 and 1.5 replicate the first four columns but are estimated using SOEs instead of private firms. While SOEs tend to have significantly better access to credit markets (as in Table 1.3), the results in these columns of Table 1.5 show that the negative correlation between sales gap and interest rates is smaller in magnitude for the SOEs than that for the private firms. Furthermore, Table 1.4 reveals that SOEs’ $D/K$ ratios do not significantly correlate with the sectoral sales gap, unlike the private firms. This finding rules out the story that the correlation between sales gap and $D/K$ ratio for private firms in each sector is driven by unobserved sectoral characteristics that determine firms’ reliance on external debt, thus lending further credibility to our interpretation that more external loans are being directed to private firms in sectors with higher sales gaps.

The Sectoral Presence of SOEs  I next examine the sectoral presence of SOEs. While they are not explicitly present in the theoretical model, SOEs can be seen as an indirect vehicle for the state to subsidize production when direct production subsidies are difficult to implement due to practical obstacles. A positive correlation between the presence of SOEs and the sectoral sales gap is therefore consistent with the hypothesis that SOEs have been placed strategically to expand sectoral production. I capture the presence of SOEs via the share of an industry’s total wage payment that is contributed by SOEs:

$$SOEshr_i \equiv \frac{\text{Total wages paid by SOEs in industry } i}{\text{Total wages paid by all firms in industry } i}.$$ (1.45)
Table 1.4: Private firms in sectors with high sales gaps receive more external loans

<table>
<thead>
<tr>
<th>Outcome Variable: Debt-To-Capital Ratio</th>
<th>Sample: Private Firms</th>
<th>Sample: SOEs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>( \hat{\xi}_i )</td>
<td>0.0711***</td>
<td>0.0705***</td>
</tr>
<tr>
<td></td>
<td>(0.0225)</td>
<td>(0.0233)</td>
</tr>
<tr>
<td>( (\bar{\varphi}_i^K)^{-1} )</td>
<td>0.00779</td>
<td>0.00847</td>
</tr>
<tr>
<td></td>
<td>(0.0258)</td>
<td>(0.0263)</td>
</tr>
<tr>
<td>( (K/Y)_i (\nu) )</td>
<td>0.0106***</td>
<td>0.0112***</td>
</tr>
<tr>
<td></td>
<td>(0.00311)</td>
<td>(0.00305)</td>
</tr>
<tr>
<td>CapForm(_{i} )</td>
<td>-0.0388**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0172)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.465***</td>
<td>0.455***</td>
</tr>
<tr>
<td></td>
<td>(0.0288)</td>
<td>(0.0400)</td>
</tr>
<tr>
<td>Obs.</td>
<td>282126</td>
<td>282126</td>
</tr>
<tr>
<td>adj. ( R^2 )</td>
<td>0.003</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Notes: The table examines the correlation between firm-level Debt-to-Capital ratio and sectoral sales gap. Columns (1) through (4) are based on the sample of private firms while columns (5) through (8) are based on the sample of SOEs. \( \hat{\xi}_i \) is the sales gap measure, as in equation (1.39). \( (\bar{\varphi}_i^K)^{-1} \) is the sectoral private return to capital inputs, as defined in equation (1.37). \( (K/Y)_i (\nu) \) is the firm-level capital intensity, defined as the ratio between capital stock and firm revenue. CapForm\(_{i} \) is the fraction of sectoral output that is unused in the accounting year and is to be used at a future time. All specifications drop outlier firms with Debt-to-Capital ratios that are either negative or above the 99th percentile. Columns (3), (4), (7), and (8) also drop outlier firms with capital intensities that are either negative or above the 99th percentile. Standard errors in parentheses are clustered at the industry level.
Table 1.5: Private firms in sectors with higher sales gaps pay lower interest rates

<table>
<thead>
<tr>
<th></th>
<th>Sample: Private Firms</th>
<th></th>
<th>Sample: SOEs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>$\xi_t$</td>
<td>-2.824***</td>
<td>-2.890***</td>
<td>-2.900***</td>
<td>-2.471***</td>
</tr>
<tr>
<td></td>
<td>(0.817)</td>
<td>(0.814)</td>
<td>(0.789)</td>
<td>(0.806)</td>
</tr>
<tr>
<td>$(\varphi_i)^{K-1}$</td>
<td>0.832</td>
<td>0.733</td>
<td>0.765</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.788)</td>
<td>(0.789)</td>
<td>(0.809)</td>
<td></td>
</tr>
<tr>
<td>$(K/Y)_i(\nu)$</td>
<td>-1.117***</td>
<td>-1.110***</td>
<td></td>
<td>-0.0872***</td>
</tr>
<tr>
<td></td>
<td>(0.179)</td>
<td>(0.178)</td>
<td></td>
<td>(0.0153)</td>
</tr>
<tr>
<td>$CapForm_i$</td>
<td>-0.785</td>
<td></td>
<td>-1.407***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.593)</td>
<td></td>
<td>(0.247)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>7.384***</td>
<td>6.335***</td>
<td>7.470**</td>
<td>7.013***</td>
</tr>
<tr>
<td></td>
<td>(1.043)</td>
<td>(1.030)</td>
<td>(1.077)</td>
<td>(1.136)</td>
</tr>
<tr>
<td></td>
<td>4.136***</td>
<td>3.041***</td>
<td>3.317***</td>
<td>2.475***</td>
</tr>
<tr>
<td></td>
<td>(0.364)</td>
<td>(0.380)</td>
<td>(0.373)</td>
<td>(0.295)</td>
</tr>
<tr>
<td>Obs.</td>
<td>163783</td>
<td>163783</td>
<td>162143</td>
<td>162143</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.006</td>
<td>0.007</td>
<td>0.026</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>0.020</td>
<td>0.024</td>
<td>0.032</td>
<td>0.039</td>
</tr>
</tbody>
</table>

Notes: The table examines the correlation between firm-level interest rates (as defined in 1.44) and sectoral sales gaps. Columns (1) through (4) are based on the sample of private firms while columns (5) through (8) are based on the sample of SOEs. $\xi_t$ is the sales gap measure, as in equation (1.39). $(\varphi_i)^{K-1}$ is the sectoral private return to capital inputs, as defined in equation (1.37). $(K/Y)_i(\nu)$ is the firm-level capital intensity, defined as the ratio between capital stock and firm revenue. $CapForm_i$ is the fraction of sectoral output that is unused in the accounting year and is to be used at a future time. All specifications drop firms with either zero interest payment or interest rates that are above the 99th percentile. Columns (3), (4), (7), and (8) also drop outlier firms with capital intensities that are either negative or above the 99th percentile. Standard errors in parentheses are clustered at the industry level. Appendix Table 1.11.4 replicates results in this table without dropping firms with zero interest payments.
I refer to $SOEshr$ as the value-added share of SOEs because labor is the only source of net value-added in the model. In Appendix Table I show that our results are robust to using SOE’s share of gross value-added (which includes wage payments as well as capital depreciation and variable profits) or total revenue to capture SOE presence.

Figure 1-3: Industries with higher sales gaps have more SOEs

Figure 1-3 plots the relationship between $SOEshr$ and sales gap for the 66 manufacturing industries, in which the size of each point in the figure reflects the total value-added of each industry. The fitted line has a significantly positive slope: SOEs constitute a higher share of industry value-added when industries have higher sales gaps. The positive relationship is robust to using SOEs’ revenue share as the outcome variable. Table 1.6 illustrates the same relationship between $SOEshr$ and sales gap using linear regressions. Column (1) represents the same information as in figure 1-3, with the positive coefficient on sales gap being the slope of the fitted line in the figure. That is, sectors with one percentage point higher sales gaps tend to have 0.418 percentage point higher SOE share of sectoral value-added. Column (2) examines the relationship between $SOEshr$ and $(\bar{\phi}_K)^{-1}$, the sectoral private return to capital goods, while column (3) regresses $SOEshr$ on both $\hat{\xi}$ and $(\bar{\phi}_K)^{-1}$. These two specifications show that $SOEshr$ is largely uncorrelated with the sectoral private return and that its correlation with the sales gap $\hat{\xi}$ remains significant and unaffected after controlling for $(\bar{\phi}_K)^{-1}$. These results imply that SOEs do not have greater presence in sectors that are themselves very constrained, lending empirical support to the theory that it is indeed the
network inefficiency captured by the sufficient statistic \( \hat{\xi} \) rather than the within-sector inefficiency that determines SOE presence. These findings echo the results in Tables 1.4 and 1.5.

Columns (4) and (5) of Table 1.6 refine the regression specification in column (3) by progressively adding sector-level control variables. The coefficient on sales gap remains largely unchanged in column (4), which controls for the capital intensity averaged across firms in the sector. Column (5) controls for \( \text{CapForm}_i \), the fraction of sectoral output that is used to form future capital stock. While the coefficient on sales gap becomes smaller in magnitude for this specification, it remains marginally significant with p-value less than 10\%, and it is statistically indistinguishable from the coefficients in columns (1) through (4). Recall that \( \text{CapForm} \) is the measure based on which we categorize industries into capital goods producers and material goods makers, and in columns (6) and (7) I investigate which of these two industry groups drives the positive relationship between sectoral SOE presence and sales gap measure. Column (6) performs the specification in column (5) on the subsample of industries that produce material goods, and column (7) does the same on the capital goods industries. The results in these specifications show that the correlation between \( \text{SOEshr} \) and the sales gap \( \hat{\xi} \) is not driven by the set of capital goods producers; instead, it is driven by the variation of sales gap within the group of material goods producers. These are industries whose output is not directly subject to credit constraints, and the variation in their sales gap is solely driven by the indirect network effect of financial frictions.

Taken together, my results in this section show that distortions in input-output linkages among the Chinese manufacturing sectors are quantitatively important, as the sales gap causes the social return to capital to be on average 12 percentage points higher than the private return. I find that private firms in sectors with higher sales gaps tend to receive more external loans and pay lower interest rates, and that Chinese SOEs are heavily directed towards sectors with larger sales gaps rather than sectors that are most constrained or have the highest private return to capital. My results suggest these interventions can be potentially welfare-enhancing because such policies effectively subsidize upstream industries (or those with large sales gaps) and as a result, these policies could address pecuniary externalities and ameliorate inefficiencies due to credit constraints faced by downstream producers. While I do not argue that my model captures the decision-making process of Chinese policymakers, my findings do allow for a positive reappraisal of the selective state interventions in the Chinese manufacturing sectors and provide a counterpoint to the prevailing view (e.g. Song et al 2011) that SOEs are a sign of sectoral inefficiency.

1.4 Cross Country Analysis

Each of the industries in this combined [input-output] table has its own peculiar input requirements, characteristic of that industry not only in the United States and in Europe but also wherever it happens to be in operation. The recipe for satisfying the appetite
Table 1.6: SOEs have higher value-added shares in industries with high sales gaps but not in those that are constrained or have high private returns to spending on capital.

<table>
<thead>
<tr>
<th>SOE share of industry value-added</th>
<th>All industries</th>
<th>Producers of material goods</th>
<th>Producers of capital goods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>(\hat{\xi}_i)</td>
<td>0.418***</td>
<td>0.417***</td>
<td>0.417***</td>
</tr>
<tr>
<td></td>
<td>(0.143)</td>
<td>(0.145)</td>
<td>(0.145)</td>
</tr>
<tr>
<td>(\left(\frac{\hat{\varphi}_K}{K/Y}\right)^{-1})</td>
<td>0.0373</td>
<td>0.00671</td>
<td>-0.0114</td>
</tr>
<tr>
<td></td>
<td>(0.106)</td>
<td>(0.101)</td>
<td>(0.103)</td>
</tr>
<tr>
<td>(\frac{K/Y}{Y}_i)</td>
<td>-0.00133</td>
<td>-0.00136</td>
<td>0.196*</td>
</tr>
<tr>
<td></td>
<td>(0.00142)</td>
<td>(0.00141)</td>
<td>(0.00142)</td>
</tr>
<tr>
<td>CapForm(_i)</td>
<td>0.211</td>
<td>-4.315</td>
<td>0.641**</td>
</tr>
<tr>
<td></td>
<td>(0.171)</td>
<td>(8.136)</td>
<td>(0.252)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.236</td>
<td>0.168</td>
<td>-0.243</td>
</tr>
<tr>
<td></td>
<td>(0.157)</td>
<td>(0.144)</td>
<td>(0.198)</td>
</tr>
<tr>
<td>Obs.</td>
<td>66</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>adj. (R^2)</td>
<td>0.104</td>
<td>-0.014</td>
<td>0.090</td>
</tr>
</tbody>
</table>

Notes: The table examines the correlation between sector-level SOE share (as defined in 1.45) and the sectoral sales gap. Columns (1) through (5) are based on the sample of all 66 sectors; column (6) is based on the subsample of 46 sectors that produce material goods; column (7) is based on the subsample of 20 sectors that produce capital goods. \(\hat{\xi}_i\) is the sales gap measure, as in equation (1.39). \(\left(\frac{\hat{\varphi}_K}{K/Y}\right)^{-1}\) is the sectoral private return to capital inputs, as defined in equation (1.37). \(\frac{K/Y}{Y}_i\) is the average capital intensity of firms in the sector. CapForm\(_i\) is the fraction of sectoral output that is unused in the accounting year and is to be used at a future time. Appendix Table 1.11.5 replicates these results by using alternative measures of SOE share.

of a blast furnace, a cement kiln, or a thermoelectric power station will be the same in India or Peru as it is, say, in Italy or California. In a sense the input-coefficient matrix derived from the U.S.-European input-output table represents a complete cookbook of modern technology. It constitutes, without doubt, the structure of a fully developed economy insofar as development has proceeded anywhere today. An underdeveloped economy can now be defined as underdeveloped to the extend that it lacks the working parts of this system.


I now turn to the second empirical exercise, in which I construct the sales gap measure for a set of developing countries using a panel of cross-country input-output tables. I then examine the
country-level correlation between the sales gap and a proxy measure of sectoral production tax rates.

1.4.1 Data Source

The main data source for my cross-country analysis is the national IO tables from OECD TiVA data set, which includes the national IO tables of 60 countries from 1995 to 2011. For the purpose of this analysis, I categorize countries into three groups according to IMF’s World Economic Outlook (IMF 2015): 24 more developed countries (MDC), 13 graduated developing countries (GDC), and 23 less developed countries (LDC). I sometimes refer to the combined MDCs and GDCs as developed countries. Table 1.11 provides a list of countries in the data set.

The input-output tables in the data are based on each individual country’s Supply and Use Table (SUT), which measures the flow of goods among industries, typically at a relatively disaggregated level (for example, the previous structural exercise with Chinese data is based on an input-output table with 135 sectors). Despite the fact that most of these countries’ statistical departments notionally follow United Nations System of National Accounts (SNA) guidelines, in practice the accounting standards and practices vary across countries, so their SUTs and IO tables are not directly comparable. Researchers on the OECD TiVA project took on the painstaking job of harmonizing the national SUTs into IO tables that are comparable across countries with 33 aggregated sectors, listed in Table 1.12, of which 15 are manufacturing industries. By my best judgement, the mapping of sectors in the data set remains imperfect. The main problem comes with the sectors “renting of machinery and equipment” and “machinery and equipment”; in some countries the former sector’s output is an order of magnitude smaller than the latter, while the pattern reverses for other countries. It is only after these two sectors are merged that their total combined output becomes comparable across countries. For this reason, I merge these two sectors label the combination “machinery and equipment” for my analysis.

There are two other data repositories for cross-country IO tables, namely the World Input-Output Database (WIOD) and Global Trade Analysis Project (GTAP). WIOD is constructed in a similar way as the OECD data, starting from the national SUTs and harmonized to match national account statistics. I do not use WIOD because it includes few developing countries, which are the focus of my analysis. The GTAP, on the other hand, is a primarily a database on bilateral trade information. Although in principle one can extract IO tables from GTAP under a more disaggregated industrial classification and for a wider range of countries, the data are constructed primarily for a different purpose and rely more heavily on imputation. As a result, the data quality of the IO tables extracted from GTAP is lower than the OECD data. For this reason, I do not use GTAP in this analysis.
1.4.2 Sales Gap

I take this section's opening quotation from Leontief (1963) seriously and infer the elasticity matrix and sales gap measure for the set of developing countries based on the observed input-output tables from the set of developed countries. Recall $\Sigma$ denotes the input-output elasticity matrix, $\Omega$ denotes the observed input-output table, and $\beta$ denotes the vector of final shares defined as (1.42). In what follows, I use subscript $c$ to refer to countries, $t$ to refer to years, and $i, j$ to refer to sectors.

**Assumption 1.4.** I make the following technology assumptions throughout this section.

a) Firms have Cobb-Douglas production technologies such that $\beta$ and $\Sigma$ are exogenous and do not change in response to allocations.

b) Firms in developed countries (MDCs and GDCs) are unconstrained and $\Sigma_{ct} = \Omega_{ct}$ for all $t$ and $c \in \text{MDC} \cup \text{GDC}$.

c) For all $t$ and $c \in \text{MDC} \cup \text{GDC}$, $\Sigma$ is a function of observable country characteristics $X_{ct}$, with entries $\sigma_{ct,ij} (X_{ct}; \theta)$ parametrized by $\theta$.

Under Assumption 1.4, I can predict the unconstrained input-output table of the developing countries using their country characteristics and the observed $\Sigma$'s for MDCs and GDCs. I implement the strategy by regressing entries of the Leontief inverse of $\Sigma_{ct}$ on the log-population for the set of developed countries, allowing for entry-year-specific constants and slopes. Specifically, I perform

$$(I - \Sigma_{ct})^{-1}_{ij} = \theta_0^{t,ij} + \theta_1^{t,ij} \cdot \ln \text{Pop}_{ct} + \epsilon_{ct,ij}$$

on $c \in \text{MDC} \cup \text{GDC}$ to form estimates $\hat{\theta}_0^{t,ij}$ and $\hat{\theta}_1^{t,ij}$. I then construct

$$(I - \hat{\Sigma}_{ct})^{-1}_{ij} = \hat{\theta}_0^{t,ij} + \hat{\theta}_1^{t,ij} \cdot \ln \text{Pop}_{ct}$$

(1.47)

for $c \in \text{LDC}$. I impute the influence vector $\mu$ and sales gap measure $\xi$ as

$$\hat{\mu}_{ct} = \frac{\beta'_c (I - \hat{\Sigma}_{ct})^{-1}}{\beta'_c (I - \Sigma_{ct})^{-1} \cdot \sigma_{ct}}$$

(1.48)

$$\hat{\xi}_{ct,ij} \equiv \frac{\hat{\mu}_{ct,ij}}{\gamma_{ct,ij}}$$

(1.49)

where $\gamma_{ct}$ the sales vector computed according to (1.40) and $\beta'_c$ again reflects the vector of consumptions shares as in (1.42). For completeness, I also construct $\hat{\xi}$ for $c \in \text{MDC} \cup \text{GDC}$ according to (1.47), (1.48), and (1.49). This imputed measure reflects differences in the predicted and actual input-output tables for these countries and it is used as a placebo check for some of my empirical
Table 1.7: Top and Bottom 5 Manufacturing Industries Ranked by Sales Gap in China

<table>
<thead>
<tr>
<th>Top 5 Sales Gap</th>
<th>Bottom 5 Sales Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fabricated metal products</td>
<td>Wood products</td>
</tr>
<tr>
<td>Misc. transport equipment</td>
<td>Rubber and plastic products</td>
</tr>
<tr>
<td>Food, beverages, and tobacco</td>
<td>Coke and petroleum products</td>
</tr>
<tr>
<td>Motor vehicles, trailers</td>
<td>Electrical machinery and apparatus</td>
</tr>
<tr>
<td>Machinery and equipment</td>
<td>Textiles, leather, footwear</td>
</tr>
</tbody>
</table>

Notes: The table ranks the 15 manufacturing sectors in the OECD data by their sectoral sales gaps in China and shows the top and bottom 5 sectors.

Population is included as a control variable since the size of an economy could determine its underlying production technologies, and I conservatively use population as the only characteristic to predict the input-output technology in order to avoid adopting variables that correlate with and potentially affect financial development or income levels. My results in this section are robust to either using population density rather than levels as the predictor or using only the year-entry fixed effects $\theta_{t,ij}$ to predict the Leontief inverse in equations (1.46) and (1.47). The results are also robust to the alternative assumption that only firms in MDCs (but not GDCs) are unconstrained.

The assumption that undistorted production technologies can be imputed based on data from developed countries is a popular approach adopted by other studies that conduct cross-country comparison of industries, including Rajan and Zingales (1998) and Hsieh and Klenow (2009). In a study more related to mine, Bartelme and Gorodnichenko (2015) also assume that wealthy nations’ IO tables represent undistorted technologies and impute that for developing countries in some of their analysis.

I now discuss the reduced form patterns of sales gap across countries. First as a validation check, in table 1.7 I list the 5 top and bottom industries, ranked by sales gap, among the 15 manufacturing industries in the OECD data for China in 2007. On the left side, with the exception of “food, beverages, and tobacco”, a puzzling outlier, the industries with high sales gaps belong to the heavy manufacturing category, which produces either capital goods or materials used for capital goods production, and it is the same set of industries that are found to have high sales gaps in the previous structural exercise based on micro data. The set of manufacturing industries that have low sales gap, which are listed on the right side of the table, consists mostly of light manufacturing industries (with the exception of the “electrical machinery and apparatus” industry). Despite the fact that these two exercises are performed under very different set of assumptions and on distinct data at different levels of aggregation, there is a consistent broad pattern—that heavy manufacturing industries in China have higher sales gaps than light manufacturing industries—which lends credibility to the empirical results.
Figure 1-4 plots the log-sales gap by sector averaged for the LDCs for year 2011, the latest year in the OECD data. The sectors are labeled on the horizontal axis and arranged according to the underlying ISIC rev 4 industry codes, from primary and light manufacturing industries on the left to heavy manufacturing and tertiary or service industries on the right. A taller bar on the figure represents a higher sales gap, with a bar above (or below) zero indicating that the industry size is too small (or large) relative to its potential sales. The broad pattern in this figure reveals that as a group, developing countries tend to have lower sales gaps in primary and light secondary industries, including agriculture, mining, and the manufacturing of consumer goods such as “textiles, leather, and footwear”, “basic chemicals”, and “non-metallic mineral products”. The LDCs tend to have higher sales gaps in the tertiary sectors and heavy manufacturing industries, such as those that produce “fabricated metal products” and “machinery and equipment”, as well as businesses that provide computer services (which includes software companies) and businesses that support R&D and business activities (including legal and consulting companies).

Independent of the theoretical importance of sales gap in the model, the pattern in figure 1-4 is striking because both influence and sales measures are constructed using each country’s own final shares $\beta_{ct}$. The difference between sectoral influence and sales, encapsulated in the sales gap measure shown in the figure, reflects only the differences between the observed input-output tables and the ones predicted based on developed countries’ IO tables. It is well known that richer countries tend to have larger tertiary sectors while poorer countries tend to have a bigger share of their economies in primary sectors and light industries, but the conventional wisdom is that such differences in relative sectoral size across countries represent either variations in specialization or non-homothetic preferences, both of which boil down to differences in the final shares across countries (Comin et al. 2016, Buera and Kaboski 2012). This figure shows that this argument is incomplete: some variation in relative sectoral size across countries with different income levels reflects differences in their underlying input-output linkages across sectors.

### 1.4.3 Sectoral Production Subsidies

Even though industrial policies are pervasive in developing countries, they are difficult to quantify with data because they almost always are implemented via a multiple explicit and implicit policy supports on many margins that could affect sectoral production. In this cross-country exercise, I turn to a measure of sectoral net (of subsidies) production taxes, recorded as part of the national input-output tables, which are available for 53 of the 60 countries in the dataset. These net taxes are net transfers to fiscal authorities incurred during production, and broadly include all miscellaneous indirect taxes and subsidies that are not commodity taxes (i.e. sales tax and value-added tax). According to the System of National Accounting (United Nations Department of Economic and Social Affairs, 1999), the national account construction standard that most countries in the data
Figure 1-4: Average Sectoral Log(Sales Gap) For LDCs
notionally follow, these net taxes should include payroll taxes (and subsidies), stamp duties, taxes for business licenses, taxes on energy use and the use of automobiles, property taxes, pollution taxes, and any monetary grant paid by government agencies to private businesses, etc. The tax rate measure created from this data is indeed a noisy measure for two reasons. First, the tax data do not distinguish among the exact margins on which the taxes or subsidies are levied, and nor do I model some of the finer details of real-world production, such as business licensing. Second, the exact underlying taxes that are aggregated to this measure differ across countries not only because tax systems are not completely comparable, particularly among developing countries, but also because countries follow the SNA guidelines to varying degrees. One major outlier is China, whose ratio of taxes to value-added is significantly higher than that of other countries. The reason is that, unlike that of other countries, the Chinese SUT compilation aggregates all indirect net taxes to this measure, including sectoral value-added tax, sales tax, and business tax (which is sales tax paid by businesses rather than consumers) in addition to the taxes outlined by the SNA standard. In the analysis below, I abstract away from the heterogeneous channels through which sectoral policies can be implemented. Instead, I take a naive approach and map the data to the model by simply assuming all taxes recorded in this variable are levied on labor, the source of net value-added in the model. The measure of sectoral subsidies is constructed according to

\[ 1 + \tau^L_{ct,i} = \frac{\text{Wage Payment}_{ct,i}}{\text{Wage Payment}_{ct,i} + \text{Net Producton Taxes}_{ct,i}}. \] (1.50)

The measure \( 1 + \tau^L_{ct,i} \) is constructed such that \( 1 + \tau^L_{ct,i} > 1 \) represents a subsidy and \( < 1 \) represents a tax. In what follows, the words “subsidy” and “tax” are used interchangeably, thereby recognizing that one is the inverse of the other. Results are reported based on dropping 1% tail on either end of the subsidy measure, and the results are robust to winsorizing instead of trimming the tails. To partially address the measurement error induced by varied accounting practices, all of the reported regression results in this section conservatively include country-by-year fixed effects, thus purging systematic differences in accounting standards across countries and time. My results are robust to excluding China.

### 1.4.4 Results

Recall again that the sales gap measure captures the ratio between social and private marginal return to expenditure on production inputs, and if network inefficiencies are of concern to fiscal authorities when designing tax policies, sectors with high sales gap should have higher subsidies \( 1 + \tau^L_i \) (or lower taxes). To check whether this pattern holds true in the data, we perform regressions of the form

\[ \log \left( 1 + \tau^L_{ct,i} \right) = \eta_0 \cdot \log \left( \xi_{ct,i} \right) + \xi_{ct} + \delta_i + \epsilon_{ct,i}, \] (1.51)
where I regress log production subsidies on log sales gap for different subsamples of countries, controlling for a full set of sector fixed effects as well as country-by-year interacted fixed effects. The sector fixed effects are introduced to eliminate the sectoral characteristics that could otherwise create variations in tax rates absent network inefficiencies; for example, when certain sectors are taxed more heavily due to pollution and other externalities. The country-by-year fixed effects are used to purge systematic differences in tax rates across countries and time, reflecting not only cross-country and temporal variation in fiscal capacity and tax optimality but also the varied accounting standards. Introducing country-by-year interacted fixed effects does not affect the interpretation because my theory suggests that if the subsidies are indeed levied on net value-added margin and are rebated back to consumers as a lump-sum transfer, multiplying subsidies across all sectors by the same constant (which translates into adding a constant in logs) does not affect allocations and it is only the relative subsidies across sectors that matter for a given country at a given time.

I perform regression (1.51) on the sample of 27 sectors rather than the full set of 33 sectors. I exclude the public sectors “public admin and defense; social security”, “education”, and “health and social work” because these are less relevant to my theory. I exclude the “agriculture” and “coke, refined petroleum products and nuclear fuel” sectors because the wage payments recorded in these two sectors are orders of magnitude too small relative to sectoral output for a number of countries and years, likely due to different accounting practices in these sectors relative to others, and, consequently, a significant number of observations in these two sectors lie outside of the 1% tail in the subsidy measure and get dropped as the tails are trimmed. My results are robust to include any or all five of the omitted sectors.

My results are shown in Table 1.8 with standard errors clustered conservatively at the country level, recognizing the potential correlations of estimation residuals within each country across sectors and time. Column (1) estimates regression (1.51) on the sample of developing countries (LDCs). The coefficient on the sales gap measures is positive and highly significant, thereby indicating that, on average and across the set of developing countries in my sample, sectors with higher sales gap tend to have higher subsidies or lower tax rates, which is consistent with the interpretation that when designing tax policies, fiscal authorities in some of these countries recognize the importance of network inefficiencies that distort the relative sector size. Column (2) estimates (1.51) based on the sample of countries in GDC \cup MDC, for which the imputed sales gap measure reflects differences between the predicted and actual input-output tables for these countries. This specification is performed as a placebo check: if firms in developed countries are truly unconstrained, there is no reason for their sectoral production tax rates to correlate with \( \log \left( \hat{\xi} \right) \), which can be interpreted as estimation error. Indeed, the coefficient on \( \log \left( \hat{\xi} \right) \) is much closer to and statistically indistinguishable from zero.

Columns (3) through (5) report results from estimating (1.51) using an alternative measure of sales gap \( \hat{\xi}^{MDC} \), calculated by modifying Assumption 1.4 and assuming that only firms in More Developed Countries are unconstrained. Specifically, the measure is constructed by estimating
equation (1.46) using only MDCs and excluding GDCs such as Hong Kong, Singapore, South Korea, and Taiwan. These three columns are then estimated on the sample of LDCs, GDCs, and MDCs, respectively. The coefficient on $\log(\tilde{\xi}_{MDC})$ for the LDC sample is significant and statistically indistinguishable from the coefficient in column (1), while those in columns (4) and (5) are indistinguishable from zero. Note that these close-to-zero coefficients are not mechanical: for example, results in column (4) imply that the sectoral tax rates in GDCs do not correlate with the sales gap measures based on differences between their IO tables and the IO tables of the MDCs, thereby indicating that financial constraints are not a first-order concern in designing tax policies in the GDCs.

Table 1.8: On average, LDCs have higher subsidies in sectors with larger sales gap

<table>
<thead>
<tr>
<th></th>
<th>Log $\left(1 + \tau^L\right)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LDC</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>$\log(\tilde{\xi})$</td>
<td>0.0125** 0.00154</td>
</tr>
<tr>
<td>$\log(\tilde{\xi}_{MDC})$</td>
<td></td>
</tr>
<tr>
<td>Obs.</td>
<td>8278</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.557</td>
</tr>
</tbody>
</table>

Notes: The table examines the correlation between sectoral sales gap and sectoral net production subsidies, as defined in (1.50), for a panel of countries between 1995 and 2011. Columns (1) and (3) are based on the sample of less developed countries; column (2) is based on the sample of graduated developing countries and more developed countries; column (4) is based on the sample of graduated developing countries; lastly, column (5) is based on the sample of more developed countries. $\log(\tilde{\xi})$ is the sectoral sales gap measure constructed under the assumption that firms in both GDCs and MDCs are unconstrained, whereas $\log(\tilde{\xi}_{MDC})$ is the sectoral sales gap measure constructed under the assumption that firms in MDCs are unconstrained. All specifications drop outlier sectors for which the net production subsidy measure lies below the 1st or above the 99th percentiles. All specifications include country-by-year and sector fixed effects. Standard errors in parentheses are clustered at the country level.

Next, I look within the LDCs and examine whether the correlation between sales gap and production subsidies is more prominent for any particular subset of countries. First, I break the countries into four groups: developing countries in 1) Asia, 2) Latin America, 3) Eastern Europe and the Middle East, and 4) Africa. There are 6 countries in each of the first 3 categories, and 2 countries in the African group. I then perform a regression pulling all 4 groups, using group
dummies interacted with $\log(\xi(i))$ to capture different elasticity of tax rates for each group. The exact specification is as follows:

$$
\log \left( 1 + r_{ct,i}^L \right) = \beta_1 \cdot 1 \text{(Asia)} \cdot \log(\xi_{ct,i}) \\
+ \beta_2 \cdot 1 \text{(South America)} \cdot \log(\xi_{ct,i}) \\
+ \beta_3 \cdot 1 \text{(Eastern Europe & Middle East)} \cdot \log(\xi_{ct,i}) \\
+ \beta_b \cdot 1 \text{(Africa)} \cdot \log(\xi_{ct,i}) \\
+ \alpha_{ct} + \delta_i + \epsilon_{ct,i}.
$$

The results are reported in table 1.9. The coefficient on $\log(\xi(i))$ is significantly different from zero only for the Asian developing countries, and the coefficients are actually slightly negative, though indistinguishable from zero, for the other three groups of countries. The result is striking: the set of Asian developing countries for which we have tax data includes China, Indonesia, Malaysia, Philippines, Thailand, and Vietnam, a group that as a whole has better economic performance over the past decade than the developing countries in the other groups. What is perhaps even more striking occurs when I re-estimate (1.52) and use country-specific dummies interacted with sales gaps; the coefficients and standard errors are in table 1.10. Good economic performers such as China, Vietnam, and Thailand have significantly positive coefficients while poor performers such as Turkey and Tunisia have significantly negative coefficients. The coefficient even picks up some variations within the Latin America region: Chile, a country with relatively good economic performance in the last decade, has a positive coefficient, while countries that suffered from weaker growth have negative coefficients.

There are two caveats in interpreting the results. First, the reduced form evidence presented in this section is on the slope of sectoral subsidies as a function of sales gap: for some countries, sectors with higher sales gap tend to have higher subsidies or lower taxes. This does not imply that these countries get the level of subsidies right: in fact, in order to fully address network inefficiencies, for many countries the subsidies need to be orders of magnitude higher than they currently are. This is hardly surprising, as sales gap is an imputed measure and is potentially noisy, thus creating attenuation bias in the estimation and weakening the coefficients towards zero. Moreover, as discussed in the theory section, tax implementation in the real world faces a wealth of practical constraints that limit the scope of intervention, and it is precisely these practical constraints and limitations that make my marginal intervention results valuable.

Second and most importantly, I emphasize that the production subsidy measure adopted in this cross-country analysis is only a noisy measure of actual subsidies and state interventions, and I take caution in interpreting the results too strongly or causally. Nevertheless, the consistent pattern that emerges from the results in this section is indicative of the importance of the network inefficiencies that my theory highlights, and my empirical findings are consistent with the hypothesis that gov-
Table 1.9: On average, Asian LDCs have higher subsidies in sectors with larger sales gap

<table>
<thead>
<tr>
<th></th>
<th>( \log \left( 1 + \tau^L \right) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Asia) ( \times \log(\hat{\xi}) )</td>
<td>0.0374*** (0.0118)</td>
</tr>
<tr>
<td>1 (South America) ( \times \log(\hat{\xi}) )</td>
<td>-0.00784 (0.0118)</td>
</tr>
<tr>
<td>1 (Eastern Europe &amp; Middle East) ( \times \log(\hat{\xi}) )</td>
<td>-0.00347 (0.0116)</td>
</tr>
<tr>
<td>1 (Africa) ( \times \log(\hat{\xi}) )</td>
<td>-0.00881 (0.0158)</td>
</tr>
<tr>
<td>Obs.</td>
<td>8278</td>
</tr>
<tr>
<td>adj. ( R^2 )</td>
<td>0.562</td>
</tr>
</tbody>
</table>

Notes: The table examines the correlation between the sectoral sales gap and sectoral net production subsidies, as defined in (1.50), for a set of less developed countries between 1995 and 2011. \( \log(\hat{\xi}) \) is the sectoral sales gap measure constructed under the assumption that firms in both GDCs and MDCs are unconstrained. All specifications drop outlier sectors for which the net production subsidy measure lies below the 1st or above the 99th percentile. All specifications include country-by-year and sector fixed effects. Standard errors in parentheses are clustered at the country level.

Governments in countries with strong economic performance understand the network distortions and are adopting policies to address them.

1.5 Conclusion

In this paper I construct a model of a production network in which firms purchase intermediate goods from each other in the presence of credit constraints. I show that these constraints distort input choices, reducing demand for upstream goods and creating a wedge between the influence and sales of upstream sectors. I further show that the ratio between the influence and sales, which I define as the sectoral sales gap, is a sufficient statistic for inefficiencies in a network and it captures the ratio between social and private marginal return to spending resources on production inputs and credit.

I conduct two distinct empirical exercises in which I estimate the sales gap measure and examine
its correlations with proxy measures of government interventions into the sector. In the context of China, I estimate the sales gap of manufacturing sectors based on firm-level production data. I find that private firms in sectors with higher sales gaps tend to receive more external loans and pay lower interest rates, and that the sectoral presence of Chinese SOEs is heavily directed towards sectors with larger sales gaps rather than sectors that are most constrained or have the highest private return to capital. My theory shows that these interventions can be welfare-enhancing because they effectively subsidize upstream industries (or those with large sales gaps), and, as a result, these policies could address pecuniary externalities and ameliorate inefficiencies due to credit constraints faced by downstream producers. My findings therefore allow for a positive reappraisal of China’s selective state interventions.

My second empirical exercise uses a panel of cross-country input-output tables to impute sales gaps for developing countries. I show that, for developing countries in Asia, the sectoral sales gap measure correlates with a measure of sectoral production subsidies, while the pattern is absent or even reversed in developing countries from other continents, which on average have had worse economic performances in recent years than their Asian counterparts. These results are consistent with the hypothesis that governments in countries with strong economic performance understand network distortions and are adopting policies to address them.

The model I present in this paper is static in nature and assumes credit constraints are imposed exogenously. A natural question to ask is whether constraints would persist if agents can save. In a related line of inquiry, I answer this question by studying a multi-sector growth model with inter-sectoral linkages and credit constraints. Entrepreneurs rationally make consumption and saving decisions, understanding that sector-specific capital stock is used both as a factor of production and also as a storage of value to serve as collateral for purchasing constrained inputs. My analysis suggests that endogenous saving is not sufficient for the economy to grow out of credit constraints. The economy features a unique equilibrium with many steady states or poverty traps, each with different levels of sectoral capital and output. The reason for stagnant economic development is demand externality: the return to saving in one sector depends on the future demand for its output, which in turn depends on the credit constraints in downstream sectors and the size of their capital stock. In a stagnant economy, capital stock is low across many sectors, and all sectors in the economy have low incentives to save. My model thus provides a dynamic, credit-based microfoundation of the "big push" theory of Rosenstein-Rodan (1943). Although it is commonly held that a "big push" environment requires large and sustained government-led investment across many sectors to take a country out of stagnation, I study development policy in this environment and show that temporary government intervention in the bottleneck sectors is sufficient to place the economy back on the path of development. The optimal development path might feature long periods of unbalanced growth as hypothesized by Hirschman (1958).
Table 1.10: Regression of sectoral subsidies on sales gap, by country

<table>
<thead>
<tr>
<th></th>
<th>Asia</th>
<th></th>
<th>Latin America</th>
<th></th>
<th>Eastern Europe &amp; Middle East</th>
<th></th>
<th>Africa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>China</td>
<td>0.0870***</td>
<td>(0.0108)</td>
<td>Indonesia</td>
<td>0.0160**</td>
<td>(0.00589)</td>
<td>Malaysia</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Thailand</td>
<td>0.0663***</td>
<td>(0.00561)</td>
<td>Vietnam</td>
</tr>
<tr>
<td></td>
<td>Philippines</td>
<td>0.0303***</td>
<td>(0.00579)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Latin America</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Argentina</td>
<td>-0.0229**</td>
<td>(0.00923)</td>
<td>Brazil</td>
<td>-0.0324**</td>
<td>(0.0124)</td>
<td>Chile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Colombia</td>
<td>-0.0747***</td>
<td>(0.0108)</td>
<td>Costa Rica</td>
<td>-0.00314</td>
<td>(0.00665)</td>
<td>Mexico</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eastern Europe &amp; Middle East</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bulgaria</td>
<td>-0.0133</td>
<td>(0.00814)</td>
<td>Croatia</td>
<td>-0.00352</td>
<td>(0.00879)</td>
<td>Hungary</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Poland</td>
<td>-0.0186</td>
<td>(0.0151)</td>
<td>Saudi Arabia</td>
<td>0.0121</td>
<td>(0.00709)</td>
<td>Turkey</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tunisia</td>
<td>-0.0633***</td>
<td>(0.0109)</td>
<td>South Africa</td>
<td>0.00314</td>
<td>(0.00640)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table reports coefficients from regressing sectoral net production subsidies, as defined in equation 1.50, on country-specific dummy variables interacted with sectoral sales gaps, based on a sample of developing countries between 1995 and 2011. The reported coefficients reflect country-specific slopes of net sectoral subsidies on sales gaps. All specifications drop outlier sectors for which the net production subsidy measure lies below the 1st or above the 99th percentiles. All specifications include country-by-year and sector fixed effects. Standard errors in parentheses are clustered at the country level.
Table 1.11: Countries in the OECD IO Table Dataset

<table>
<thead>
<tr>
<th>More Developed Countries (MDC)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>Finland</td>
</tr>
<tr>
<td>Austria</td>
<td>France</td>
</tr>
<tr>
<td>Belgium</td>
<td>Germany</td>
</tr>
<tr>
<td>Brunei Darussalam*</td>
<td>Greece</td>
</tr>
<tr>
<td>Canada</td>
<td>Iceland*</td>
</tr>
<tr>
<td>Denmark</td>
<td>Ireland</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Graduated Developing Countries (GDC)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hong Kong (China)</td>
<td>Israel</td>
</tr>
<tr>
<td>Cyprus*</td>
<td>Latvia</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>Lithuania</td>
</tr>
<tr>
<td>Estonia</td>
<td>Malta</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Less Developed Countries (LDC)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>Colombia</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>Costa Rica</td>
</tr>
<tr>
<td>Brazil</td>
<td>Croatia</td>
</tr>
<tr>
<td>Cambodia*</td>
<td>Hungary</td>
</tr>
<tr>
<td>Chile</td>
<td>India*</td>
</tr>
<tr>
<td>China</td>
<td>Indonesia</td>
</tr>
</tbody>
</table>

* There are no data on production taxes for these countries.
Table 1.12: Non-Public Sectors in the OECD IO Table Dataset

<table>
<thead>
<tr>
<th>33 Sectors in the OECD IO Tables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
</tr>
<tr>
<td>Mining</td>
</tr>
<tr>
<td>Food products, beverages, and tobacco</td>
</tr>
<tr>
<td>Textiles, leather, footwear</td>
</tr>
<tr>
<td>Wood products</td>
</tr>
<tr>
<td>Pulp, paper products, printing</td>
</tr>
<tr>
<td>Coke, refined petroleum products and nuclear fuel</td>
</tr>
<tr>
<td>Chemicals and chemical products</td>
</tr>
<tr>
<td>Rubber and plastic products</td>
</tr>
<tr>
<td>Other non-metallic mineral products</td>
</tr>
<tr>
<td>Basic metals</td>
</tr>
<tr>
<td>Fabricated metal products</td>
</tr>
<tr>
<td>Machinery and equipment</td>
</tr>
<tr>
<td>Computer, electronic and optical equipment</td>
</tr>
<tr>
<td>Electrical machinery and apparatus</td>
</tr>
<tr>
<td>Motor vehicles, trailers</td>
</tr>
<tr>
<td>Other transport equipment</td>
</tr>
<tr>
<td>Recycling</td>
</tr>
<tr>
<td>Electricity, gas and water supply</td>
</tr>
<tr>
<td>Construction</td>
</tr>
<tr>
<td>Wholesale and retail trade</td>
</tr>
<tr>
<td>Hotels and restaurants</td>
</tr>
<tr>
<td>Transport and storage</td>
</tr>
<tr>
<td>Post and telecommunications</td>
</tr>
<tr>
<td>Financial intermediation</td>
</tr>
<tr>
<td>Real estate activities</td>
</tr>
<tr>
<td>Renting of machinery and equipment*</td>
</tr>
<tr>
<td>Computer and related activities</td>
</tr>
<tr>
<td>R&amp;D and business activities</td>
</tr>
<tr>
<td>Public admin and defense; social security</td>
</tr>
<tr>
<td>Education</td>
</tr>
<tr>
<td>Health and social work</td>
</tr>
<tr>
<td>Other community, social, and personal service</td>
</tr>
</tbody>
</table>

* This sector is merged with "Machinery and equipment" sector for all of the analysis in this section.
1.6 Appendix: Proofs

In this section of the appendix I provide proofs for results in the model without Cobb-Douglas assumptions. Proof of Theorem 1.3 is deferred to Appendix 1.7.2, after we provide closed form solutions of the equilibrium for the Cobb-Douglas case in Appendix 1.7.1.

I use boldface Roman alphabets and greek letters without subscripts to represent column vectors (e.g. \( \mathbf{p} \equiv (p_1, \cdots, p_S)^T \) and \( \mathbf{\theta} \equiv (\theta_1, \cdots, \theta_S)^T \)), and let \( \mathbf{1} \) be a column vector of ones.

1.6.1 Proof of Proposition 1.2

Let \( \pi_i (\nu; C_i, w, \{p_j\}) \) be the solution to the profit maximization problem (Pfirm) for firm \( \nu \) in industry \( i \), where \( C_i \) is the price at which firm \( \nu \) sells its output and \( w \) and \( \{p_j\} \) are the cost of inputs. Note that \( C_i = p_i \) in equilibrium, but the sectoral unit cost \( C_i (\cdot) \) can be defined as a function of \( p_i \) as well as the prices of other production inputs as in (1.15). Totally differentiating \( \pi_i \) and applying the Envelope theorem yields:

\[
\frac{d\pi_i (\nu)}{d\ln C_i} = C_i q_i (\nu) \left( d\ln C_i - \alpha_i (\nu) d\ln w - \sum_j \sigma_{ij} (\nu) d\ln p_j \right),
\]

which implies

\[
\frac{\partial \pi_i (\nu)}{\partial \ln C_i} = C_i q_i (\nu),
\]

\[
\frac{\partial \pi_i (\nu)}{\partial \ln w} = -\alpha_i (\nu) C_i q_i (\nu),
\]

\[
\frac{\partial \pi_i (\nu)}{\partial \ln p_j} = -\sigma_{ij} (\nu) C_i q_i (\nu).
\]

The free-entry condition defines an implicit function of \( C_i \) and the input prices \( w \) and \( \{p_j\} \):

\[
\kappa = \int \pi_i (\nu; C_i (w, \{p_j\}), w, \{p_j\}) d\Phi_i (\nu)
\]

Totally differentiating the free-entry condition, we get

\[
0 = \int q_i (\nu) \left( d\ln C_i - \alpha_i (\nu) d\ln w - \sum_j \sigma_{ij} (\nu) d\ln p_j \right) d\Phi_i (\nu)
\]

Applying the Implicit Function Theorem, we have

\[
\frac{\partial \ln C_i}{\partial \ln p_j} = \int \sigma_{ij} (\nu) \frac{q_i (\nu)}{q_i (\nu) d\Phi_i (\nu)} d\Phi_i (\nu)
\]

\[
= \mathbb{E}_\nu \left[ \sigma_{ij} (\nu) \frac{q_i (\nu)}{\mathbb{E}_\nu [q_i (\nu)]} \right].
\]
We can similarly find the elasticity of the sectoral unit-cost function with respect to the wage rate, which we will use later in this appendix:

\[ \alpha_i \equiv \mathbb{E}_\nu \left[ \alpha_i(\nu) \frac{q_i(\nu)}{\mathbb{E}_\nu[q_i(\nu)]} \right] = \frac{\partial \ln C_i}{\partial \ln w}. \tag{1.6.2} \]

### 1.6.2 Proof of Proposition 1.3

**Influence** Recall from (1.16) and (1.17) that the equilibrium price \( \{p_j\} \) vector is the fixed point to the system of equations

\[ p_i = C_i(w, \{p_j\}; h_i) \]

with the normalization

\[ 1 = CF(\{p_j\}). \]

Totally differentiating, we have

\[ d \ln p_i = -d \ln h_i + \alpha_i d \ln w + \sum \sigma_{ij} d \ln p_j \]

\[ 0 = \sum \beta_i d \ln p_i \]

where the first equation follows from Proposition 1.2 and equation (1.6.2). Re-writing these equations using matrix notation, we have

\[ d \ln p = -d \ln h + \alpha d \ln w + \Sigma d \ln p \]

\[ = (I - \Sigma)^{-1} (-d \ln h + \alpha d \ln w) \]

and

\[ 0 = \beta' d \ln p \]

where \( \alpha' \equiv (\alpha_1, \cdots, \alpha_S) \) is the vector of sectoral average labor share. The consumer budget constraint and the resource constraint implies \( wL = Y \) hence

\[ d \ln Y = d \ln w \]

\[ = \frac{\beta'(I - \Sigma)^{-1}}{\beta'(I - \Sigma)^{-1} \alpha} \cdot d \ln h \]

which proves the result.
Sales  The market clearing condition for good $j$ is

$$Q_j = Y_j + \sum_i M_{ij}$$

Multiplying by $p_j/Y$ and using the fact that $\gamma_j \equiv \frac{p_j Q_j}{Y}$ and $\beta_j \equiv \frac{p_j Y_j}{F(Y_1, \ldots, Y_S)}$, we obtain

$$\gamma_j = c \cdot \beta_j + \sum_i \omega_{ij} \gamma_i$$

or in matrix notation,

$$\gamma' = c \cdot \beta' (I - \Omega)^{-1}$$

where $c \equiv \frac{F(Y_1, \ldots, Y_S)}{Y}$ is a scalar. To figure out the value for $c$, note that the total wage payment in industry $i$ can be written as

$$wL_i = \alpha_i p_i Q_i$$

or

$$wL = \sum_i wL_i = \left( \sum_i \alpha_i \gamma_i \right) Y.$$  

The consumer budget constraint and the resource constraint imply $wL = Y$, or

$$\sum_i \gamma_i \alpha_i = 1.$$  

The constant $c$ can be therefore found as

$$c = \frac{1}{\beta' (I - \Omega)^{-1}}.$$

1.6.3 Proof of Theorem 1.1

In presence of taxes, equation (1.6.1) can be modified as

$$d\pi_i (\nu) = C_i q_i (\nu) \left( d\ln C_i - \alpha_i (\nu) d\ln w - \sum_j \sigma_{ij} (\nu) d\ln p_j \right)$$

$$+ C_i q_i (\nu) \left( d\ln \left( 1 + \tau_i^R \right) + \alpha_i (\nu) d\ln \left( 1 + \tau_i^L \right) + \sum_j \sigma_{ij} (\nu) \ln \left( 1 + \tau_i^L \right) \right).$$
Exploiting (1.6.3) and applying the same argument as in the proof for Proposition 1.2, we have

\[
\begin{cases}
\frac{\partial \ln C_i (w, \{p_i\}, \tau^R, \tau^L, \{\tau'_j\})}{\partial \ln (1 + \tau_i)} = -1, \\
\frac{\partial \ln C_i (w, \{p_i\}, \tau^R, \tau^L, \{\tau'_j\})}{\partial \ln (1 + \tau'_i)} = -\alpha_i, \\
\frac{\partial \ln C_i (w, \{p_i\}, \tau^R, \tau^L, \{\tau'_j\})}{\partial \ln (1 + \tau'_j)} = -\sigma_{ij}.
\end{cases}
\quad (1.6.4)
\]

Similar to the proof of Proposition 1.3, for each subsidy \( \tau_i \in \{\tau^R, \tau^L, \{\tau'_j\}\} \) we can totally differentiate the sectoral unit-cost functions and obtain a system of equations

\[
d \ln p_j = \alpha_j d \ln w + \sum \sigma_{jk} d \ln p_k + \frac{\partial \ln C_j}{\partial \ln (1 + \tau_i)} d \ln (1 + \tau_i)
\]

in which \( \frac{\partial \ln C_j}{\partial \ln (1 + \tau_i)} = 0 \) for all \( i \neq j \). Manipulating the equations and using matrix notations, we have

\[
\frac{d \ln w}{d \ln (1 + \tau_i)} = -\frac{\beta'}{(I - \Sigma)^{-1}} \alpha \left( \frac{\partial \ln C}{\partial \ln (1 + \tau_i)} \right)
\]

\[
= -\mu \frac{\partial \ln C_i}{\partial \ln (1 + \tau_i)}.
\]

Lastly, from the budget constraints for the consumer and the planner

\[
wL = C + T
\]

\[
T = E + \sum_{i=1}^{S} \left( \tau^R_i p_i Q_i + \tau^L_i w L_i + \sum_{j=1}^{S} \tau'_j p_j M_{ij} \right)
\]

as well as the resource constraint of the economy

\[
Y = C + E
\]

we obtain

\[
wL = Y + \sum_{i=1}^{S} \left( \tau^R_i p_i Q_i + \tau^L_i w L_i + \sum_{j=1}^{S} \tau'_j p_j M_{ij} \right)
\]

which implies

\[
\frac{d \ln Y}{d \ln (1 + \tau_i)} \bigg|_{\tau=0, \text{holding } E \text{ constant}} = \left( \frac{d \ln w}{d \ln (1 + \tau_i)} - \frac{dT/d\tau_i}{Y} \right) \bigg|_{\tau=0, \text{holding } E \text{ constant}}.
\]
Using (1.6.4) and the fact that sectoral expenditure shares follow

\[ wL_i = \alpha_i p_i Q_i, \quad p_j M_{ij} = \omega_{ij} p_i Q_i, \]

we have

\[
\begin{align*}
\frac{d \ln Y}{d \ln (1+\tau_i)} \bigg|_{\tau = 0, \text{holding } E \text{ constant}} &= \mu_i - \gamma_i \\
\frac{d \ln Y}{d \ln (1+\tau_i)} \bigg|_{\tau = 0, \text{holding } E \text{ constant}} &= \alpha_i (\mu_i - \gamma_i) \\
\frac{d \ln Y}{d \ln (1+\tau_i)} \bigg|_{\tau = 0, \text{holding } E \text{ constant}} &= \sigma_{ij} \mu_i - \omega_{ij} \gamma_i,
\end{align*}
\]

which proves parts 1, 3, and 4 of the theorem. Part 2 follows part 3 and the observation that \( \sigma_{ij} = \omega_{ij} \) for all unconstrained intermediate inputs \( j \notin K_i \).

### 1.6.4 Proof of Theorem 1.2

From the resource constraint we derive that for any subsidy \( \tau_i \in \{ \tau_i^R, \tau_i^L, \{ \tau_j \} \} \),

\[
\frac{dY}{d \ln (1+\tau_i)} = \frac{dC}{d \ln (1+\tau_i)} + \frac{dE}{d \ln (1+\tau_i)}. \tag{1.6.6}
\]

Starting from \( \tau = 0 \) and holding the lump sum tax \( T \) constant, we have

\[
\begin{align*}
\frac{dE}{d \ln (1+\tau_i^R)} \bigg|_{\tau = 0, \text{holding } T \text{ constant}} &= -\gamma_i Y \\
\frac{dE}{d \ln (1+\tau_i^L)} \bigg|_{\tau = 0, \text{holding } T \text{ constant}} &= -\alpha_i \gamma_i Y \\
\frac{dE}{d \ln (1+\tau_i^L)} \bigg|_{\tau = 0, \text{holding } T \text{ constant}} &= -\omega_{ij} \gamma_i Y.
\end{align*}
\]

Combining with (1.6.5) and (1.6.6), this implies

\[
\begin{align*}
\frac{dC}{d \ln (1+\tau_i^R)} \bigg|_{\tau = 0, \text{holding } T \text{ constant}} &= \mu_i Y \\
\frac{dC}{d \ln (1+\tau_i^L)} \bigg|_{\tau = 0, \text{holding } T \text{ constant}} &= \alpha_i \mu_i Y \\
\frac{dE}{d \ln (1+\tau_i^L)} \bigg|_{\tau = 0, \text{holding } T \text{ constant}} &= \sigma_{ij} \mu_i Y.
\end{align*}
\]
The theorem follows from these two sets of equations and the fact that \( d \ln (1 + \tau_i) \bigg|_{\tau=0} = d\tau_i. \)

1.6.5 Proof of Propositions 1.4 and 1.5

Following a similar procedure as in the proof for Proposition 1.2, we can show

\[
\frac{\partial \ln C_i}{\partial \ln (1 + \tau_i^C)} \bigg|_{\tau=0} = -\mathbb{E}_\nu \left[ \left( \varphi_i^K (\nu)^{-1} - 1 \right) \frac{W_i (\nu)}{\mathbb{E}_\nu [p_i q_i (\nu) \nu]} \right].
\]

We then use a similar procedure as in the proof for Theorem 1.1 to show

\[
\frac{d \ln w}{d \ln (1 + \tau_i^C)} \bigg|_{\tau=0} = \mu_i \times \mathbb{E}_\nu \left[ \left( \varphi_i^K (\nu)^{-1} - 1 \right) \frac{W_i (\nu)}{\mathbb{E}_\nu [p_i q_i (\nu) \nu]} \right].
\]

Lastly, differentiating (1.26) with respect to \( \tau_i^C \) and using planner’s budget constrain (1.27), we have

\[
\frac{dE}{d\tau_i^C} \bigg|_{\tau=0, \text{holding } T \text{ constant}} = -\gamma_i \chi \frac{\mathbb{E} \left[ W_i (\nu) 1 \left( \varphi_i (\nu)^{-1} > 1 \right) \right]}{\mathbb{E} [p_i q_i (\nu) \nu]}.
\]

The social return to credit \( SR_i^C \) can be found similarly as in the proof of Theorem 1.2:

\[
SR_i^C = \left. \frac{dC/d\tau_i^C}{dE/d\tau_i^C} \right|_{\tau=0, \text{holding } T \text{ constant}} = \frac{d \ln w / d \ln (1 + \tau_i^C)}{(dE/d\tau_i^C) / Y} \bigg|_{\tau=0, \text{holding } T \text{ constant}} = \xi_i \cdot \chi^{-1} \cdot \mathbb{E} \left[ \left( \varphi_i^{-1} (\nu) - 1 \right) \frac{W_i (\nu)}{\mathbb{E} [W_i (\nu) 1 \left( \varphi_i (\nu)^{-1} > 1 \right)]} \right],
\]

which proves Proposition 1.4 as we note that \( PR_i^C \equiv \chi^{-1} \cdot \mathbb{E} \left[ \left( \varphi_i^{-1} (\nu) - 1 \right) \frac{W_i (\nu)}{\mathbb{E} [W_i (\nu) 1 \left( \varphi_i (\nu)^{-1} > 1 \right)]} \right]. \)

To prove Proposition 1.5, we follow the same steps to show

\[
\frac{d \ln w}{d\tau_i^C} \bigg|_{\tau=0} = \mu_i \frac{\mathbb{E} \left[ \varphi_i (\nu)^{-1} - 1 \right]}{\mathbb{E} [p_i q_i (\nu) \nu]}
\]

and

\[
\frac{dE}{d\tau_i^C} \bigg|_{\tau=0} = -\gamma_i \chi Y Pr (\varphi_i (\nu) < 1) \frac{\mathbb{E} [p_i q_i (\nu) \nu]}{\mathbb{E} [p_i q_i (\nu) \nu]}.
\]
The social return is thus

\[
SR_i^{C'} = \left. -\frac{d\ln w}{dC_i} \right|_{T=0} \cdot \frac{dE}{dC_i} / Y = \xi_i \cdot \chi^{-1} \cdot \mathbb{E} \left[ \varphi_i^{-1} (\nu) - 1 \mid \varphi_i (\nu) < 1 \right],
\]

as desired.

### 1.6.6 Proof of Proposition 1.6 and Derivation of Equation (1.30)

From Proposition 1.3 we have

\[
\mu' \propto \beta' (I - \Sigma)^{-1}, \quad \gamma' \propto \beta' (I - \Omega)^{-1}
\]

which implies that for some constants \(c_1\) and \(c_2\),

\[
c_1\mu' - c_2\gamma' = c_1\mu'\Sigma - c_2\gamma'\Omega.
\]

Writing out the \(j\)-th equation of the system above and dividing by \(c_2\gamma_j\), we have

\[
\frac{c_1\mu_j - c_2\gamma_j}{c_2\gamma_j} = \sum_i \left( \sigma_{ij} \frac{c_1\mu_i - c_2\gamma_i}{c_2\gamma_j} + \frac{\gamma_i}{\gamma_j} (\sigma_{ij} - \omega_{ij}) \right)
\]

\[
= \sum_i \left( \delta_{ij} \frac{c_1\mu_i - c_2\gamma_i}{c_2\gamma_i} + (\delta_{ij} - \omega_{ij}) \right)
\]

where the second equality follows from \(\delta_{ij} = \omega_{ij} \frac{\gamma_i}{\gamma_j}\) and \(\delta_{ij} = \sigma_{ij} \frac{\gamma_i}{\gamma_j}\). Stacking the equations using matrix notations, we have

\[
\begin{pmatrix}
\frac{c_1}{c_2} \\
\end{pmatrix} \xi' - 1' = 1' \left( \hat{\Sigma} - \hat{\Omega} \right) \left( I - \hat{\Sigma} \right)^{-1}
\]

(1.6.7)

\[
\Rightarrow \xi' \propto 1' \left( I + \hat{\Sigma} \left( I - \hat{\Sigma} \right)^{-1} - \hat{B} \left( I - \hat{\Sigma} \right)^{-1} \right)
\]

\[
= 1' \left( I - \hat{\Omega} \right) \left( I - \hat{\Sigma} \right)^{-1}.
\]
To obtain equation (1.30), note that the assumption $\omega_{ij} = \sigma_{ij} \varphi$ implies $\hat{\Omega} = \hat{\Sigma} \varphi$ hence equation (1.6.7) implies

$$
\xi' \propto 1' + 1' \left( \hat{\Sigma} - \hat{\Omega} \right) \left( I - \hat{\Sigma} \right)^{-1} \\
= \varphi 1' + (1 - \varphi) 1'I + (1 - \varphi) 1'\hat{\Sigma} \left( I - \hat{\Sigma} \right)^{-1} \\
= \varphi + (1 - \varphi) 1' \left( I - \hat{\Sigma} \right)^{-1} \\
\propto \text{const} + \left( \delta^{\text{Jones}} \right)',
$$

1.7 Appendix: Cobb-Douglas Production Functions

1.7.1 Cobb-Douglas Fully Solved

In this subsection of the appendix I setup the model with Cobb-Douglas production function and fully solve for the equilibrium. I then prove Theorem 1.3 in the next subsection. For simplicity, I exposit without firm-level heterogeneity, which can be added with minor modifications in notations.

**Setup** Potential entrepreneurs pay fixed cost $\kappa_i$ units of the final good to acquire a production function

$$
q_i = z_i \xi_i^{\alpha_i} \prod_j m_i^{\sigma_{ij}}
$$

and working capital $W_i$ with constraint

$$
\sum_{j \in K_i} p_j m_{ij} \leq W_i.
$$

The final good is produced according to

$$
F = \prod_i Y_i^{\beta_i}.
$$

**Equilibrium** An equilibrium is the a of allocation and prices such that 1) all producers maximize profits taking prices and constraints as given, 2) free-entry drives ex-ante profits to zero, and 3) all markets clear.

Let small-case letters denote firm-level variables and let capital letters denote sectoral and aggregate variables, and let $N_i$ be the number of firms that enter sector $i$ in equilibrium. The market clearing conditions are

$$
Q_j = Y_j + \sum_i M_{ij} \text{ for all } j.
$$
\[ F = Y + \sum_i \kappa_i N_i \]

where

\[ Q_i = N_i q_i, \quad L_i = N_i \ell_i, \quad M_{ij} = N_i m_{ij}. \]

**Firm Allocations** Profit maximization implies

\[ w \ell_i = \alpha_i p_i q_i \]

\[ p_{jm_{ij}} = \begin{cases} \sigma_{ij} p_i q_i & \text{for } j \notin K_i \\ \sum_{j \in K_i} \sigma_{ij} \min \left\{ \sum_{j \in K_i} \sigma_{ij} p_i q_i, W_i \right\} & \text{for } j \in K_i. \end{cases} \]

**Sectoral Allocations** Free-entry implies

\[ \kappa_i = \left( 1 - \alpha_i - \sum_{j \notin K_i} \sigma_{ij} \right) p_i q_i - \min \left\{ \sum_{j \in K_i} \sigma_{ij} p_i q_i, W_i \right\} \]

which implies that constraints bind in sector \( i \) if and only if

\[ \frac{W_i}{\kappa_i} < \frac{1 - \alpha_i - \sum_{j \notin K_i} \sigma_{ij}}{1 - \alpha_i - \sum_j \sigma_{ij}}. \]

Let

\[ \varphi_i \equiv \min \left\{ 1, \frac{W_i}{W_i + \kappa_i} \left( 1 - \alpha_i - \sum_{j \notin K_i} \sigma_{ij} \right) \right\} \]

and we can rewrite the firm-level allocation of constrained inputs as

\[ p_{jm_{ij}} = \sigma_{ij} \varphi_i p_i q_i \quad \text{for } j \in K_i. \]

To obtain sectoral allocations, we multiply both sides of the firm-level allocation equation by \( N_i \) to get

\[ wL_i = \alpha_i p_i Q_i \]

\[ p_{jm_{ij}} = \begin{cases} \sigma_{ij} p_i Q_i & \text{for } j \notin K_i \\ \sigma_{ij} \varphi_i p_i Q_i & \text{for } j \in K_i \end{cases} \]

\[ \kappa_i N_i = \left( 1 - \alpha_i - \sum_{j \notin K_i} \sigma_{ij} - \varphi_i \sum_{j \in K_i} \sigma_{ij} \right) p_i Q_i \]

87
**Aggregate Allocations** Define $\gamma_i \equiv \frac{\partial Q_i}{\partial y}$ as the sales vector. Labor market clearing implies

$$L_i = \frac{\alpha_i \gamma_i}{\sum_j \alpha_j \gamma_j} L.$$  

Using the fact that $wL = Y$, we have $\sum wL_i = \sum \alpha_i \gamma_i Y = Y$ which implies the denominator in the equation above is 1 and that

$$L_i = \alpha_i \gamma_i L. \quad (1.7.1)$$

The sectoral allocation for intermediate goods can be re-written as

$$M_{ij} = \begin{cases} \sigma_{ij} \gamma_{ij} Q_j & \text{for } j \notin K_i \\ \sigma_{ij} \varphi_i \gamma_{ij} Q_j & \text{for } j \in K_i \end{cases}$$

We can solve for aggregate allocations by substituting sectoral allocations into production functions. The sectoral production function for sector $i$ can be aggregated from firm-level production functions as

$$Q_i = N_i \eta_i$$

$$= N_i z_i \ell_i \prod_j m_{ij}^{\sigma_{ij}}$$

$$= z_i N_i \left(1 - \alpha_i - \sum_j \sigma_{ij}\right) \prod_j M_{ij}^{\sigma_{ij}}$$

$$= z_i \left(1 - \alpha_i - \sum_{j \notin K_i} \sigma_{ij} - \varphi_i \sum_{j \in K_i} \sigma_{ij} \gamma_i Y \right) \left(1 - \alpha_i - \sum_j \sigma_{ij}\right)$$

$$\times \left(\sum_{j \notin K_i} \sigma_{ij} \gamma_{ij} Q_j \right) \prod_{j \in K_i} \varphi_i \sigma_{ij}$$

Taking logs and move $\ln \gamma_i$ to the left-hand-side,

$$\ln Q_i - \ln \gamma_i = \ln \omega_i + \ln z_i + \left(1 - \alpha_i - \sum_j \sigma_{ij}\right) \ln Y + \alpha_i \ln L + \sum \sigma_{ij} \left(\ln Q_j - \ln \gamma_j\right)$$

$$+ \left(1 - \alpha_i - \sum_j \sigma_{ij}\right) \ln \left(1 - \alpha_i - \sum_{j \notin K_i} \sigma_{ij} - \varphi_i \sum_{j \in K_i} \sigma_{ij}\right) + \sum \sigma_{ij} \ln \varphi_i$$

where $c_i \equiv \left(1 - \alpha_i - \sum_j \sigma_{ij}\right) \ln \left(1 - \alpha_i - \sum_j \sigma_{ij}\right) + \alpha_i \ln \alpha_i + \sum_j \sigma_{ij} \ln \sigma_{ij}$. To economize notation, let

$$\phi_i \equiv \left(1 - \alpha_i - \sum_{j \notin K_i} \sigma_{ij} - \varphi_i \sum_{j \in K_i} \sigma_{ij}\right)$$

$$\frac{1 - \alpha_i - \sum_j \sigma_{ij}}{1 - \alpha_i - \sum_j \sigma_{ij}}$$

88
\[ \theta_i = \left(1 - \alpha_i - \sum_j \sigma_{ij}\right) \]
\[ \sigma^K_i = \sum_{j \in K_i} \sigma_{ij}. \]

We can then write
\[ \ln Q_i - \ln \gamma_i = \ln c_i + \ln z_i + \theta_i \ln Y + \alpha_i \ln L + \sum \sigma_{ij} (\ln Q_j - \ln \gamma_j) + \theta_i \ln \phi_i + \sigma^K_i \ln \varphi_i. \] (1.7.2)

Note that when credit constraints do not bind in the sector, we have \( \phi_i = \varphi_i = 1. \)

Profit-maximization by the final good producer implies
\[ p_i Y_i = \beta_i F. \]

The sales vector can therefore be re-written as
\[ \gamma_i = \frac{p_i Q_i}{Y_i} = \frac{Q_i \beta_i F}{Y_i Y}. \]

Thus
\[ Y_i = \frac{Q_i \beta_i F}{\gamma_i Y}. \]

The net aggregate output can be found by using the production function of the final good:
\[ \ln Y = \ln F + (\ln Y - \ln F) = \beta' (\ln Q - \ln \gamma + \ln \beta + \ln F - \ln Y) + \ln Y - \ln F \]
\[ = \beta' (\ln Q - \ln \gamma + \ln \beta) \]

and from equation (1.7.2) we have
\[ \ln Q - \ln \gamma = (I - \Sigma)^{-1} \left( \ln c + \ln z + \theta \ln Y + \alpha \ln L + \theta \ln \phi + \sigma^K \ln \varphi \right). \]

Hence
\[ \ln Y = \beta' \ln \beta + \beta' (I - \Sigma)^{-1} \left( \ln c + \ln z + \theta \ln Y + \alpha \ln L + \theta \ln \phi + \sigma^K \ln \varphi \right) \]
\[ = \frac{\beta' \ln \beta + \beta' (I - \Sigma)^{-1} \left( \ln c + \ln z + \alpha \ln L + \theta \ln \phi + \sigma^K \ln \varphi \right)}{1 - \beta' (I - \Sigma)^{-1} \theta} \left( \ln c + \ln z + \alpha \ln L + \theta \ln \phi + \sigma^K \ln \varphi \right) \]

89
Given that
\[ \theta_i = 1 - \alpha_i - \sum_{j} \sigma_{ij} \]
we have
\[ \beta'(I - \Sigma)^{-1} (\theta + \alpha) = \beta'(I - \Sigma)^{-1} (1 - \Sigma 1) \]
\[ = \beta'(I - \Sigma)^{-1} (I - \Sigma) 1 \]
\[ = 1. \]

Thus we can write \( \ln Y \) as
\[ \ln Y = \frac{\beta' \left( \ln \beta + (I - \Sigma)^{-1} \ln c \right)}{\beta' (I - \Sigma)^{-1} \alpha} + \frac{\beta' (I - \Sigma)^{-1}}{\beta' (I - \Sigma)^{-1} \alpha} (\ln z + \theta \ln \phi + \sigma^K \ln \varphi) + \ln L \]

The first additive term is a scalar that depends on the Cobb-Douglas coefficients.

We can immediately see that (1) there is aggregate constant returns to scale as the net aggregate consumption is linear in \( L \); (2) \( \frac{\beta'(I - \Sigma)^{-1}}{\beta'(I - \Sigma)^{-1} \alpha} \) is the influence vector \( \mu_i \); and (3) credit constraints lower effective sectoral productivity by a factor of \( \phi_i^\theta \phi_i^\sigma \), and they lower aggregate productivity by \( \left( \phi_i^\theta \phi_i^\sigma \right)^K \): that is, the sectoral productivity shock is transmitted to affect aggregate productivity via the influence of the sector.

Once we know \( Y \), solving for sectoral allocations and prices boils down to solving for the sales vector \( \gamma \), which can be found from the market clearing conditions:

\[ p_j Q_j = p_j Y_j + \sum_i p_j M_{ij} \]

\[ \iff \gamma_j = \beta_j \frac{F}{Y} + \sum_\omega \omega_{ij} \gamma_i \]  \hspace{1cm} (1.7.3)

where
\[ \omega_{ij} = \begin{cases} \sigma_{ij} & \text{if } j \notin K_i \\ \varphi_i \sigma_{ij} & \text{otherwise.} \end{cases} \]

and the matrix \( \Omega = [\omega_{ij}] \) is exactly the input-output table. Stacking (1.7.3) into matrix notation, we have

\[ \gamma' = \frac{\beta'F}{Y} + \gamma' \Omega \]
\[ = \frac{F}{Y} \beta' (I - \Omega)^{-1} \]
\[ = \frac{\beta' (I - \Omega)^{-1}}{\beta' (I - \Omega)^{-1} \alpha} \]

90
where the third equality follows from the fact that $\gamma'\alpha = 1$ as in (1.7.1).

1.7.2 Proof of Theorem 1.3

Consider labor subsidies $\{r^L_i\}$. The sectoral allocation of labor, intermediate inputs, and number of firms can be re-written as

$$L_i = \frac{(1 + r^L_i) \gamma_i \alpha_i}{\sum_j (1 + r^L_j) \gamma_j \alpha_j} L$$  \hspace{1cm} (1.7.4)

$$M_{ij} = \begin{cases} 
\sigma_{ij} \varphi_{ij} \frac{2K}{\gamma_j} Q_j & \text{for } j \notin K_i \\
\sigma_{ij} \varphi_{ij} \frac{2K}{\gamma_j} Q_j & \text{for } j \in K_i
\end{cases}$$  \hspace{1cm} (1.7.5)

$$\kappa_i N_i = \left(1 - \alpha_i - \sum_{j \notin K_i} \sigma_{ij} - \varphi_i \sum_{j \in K_i} \sigma_{ij}\right) \gamma_i Y$$

Following the derivation in Appendix 1.7.1, we can write total output as

$$\ln Y = \text{const} + \sum \mu_i \alpha_i \ln (1 + r^L_i) - \ln \left(\sum_j (1 + r^L_j) \gamma_j \alpha_j\right)$$

The planner's problem is therefore to solve

$$\max_{\{r^L_i\}} \sum \mu_i \alpha_i \ln (1 + r^L_i) - \ln \left(\sum_j (1 + r^L_j) \gamma_j \alpha_j\right)$$

Note in particular that we do not need to impose the budget constraint for the maximization problem for two reasons. First, the lump sum tax $T$ does not affect allocations and can be chosen to satisfy the budget constraint after $\{r^L_i\}$ are pinned down as the solution to the maximization problem. Second, it should be clear from the objective function that it is the proportionality of $(1 + r^L_i)$ across sectors that matters, rather than the levels of subsidies. As a result, for any given lump sum tax $T$, the planner can always rescale the subsidies by a constant and balance his budget without affecting allocation. It is therefore without loss of generality to set $T = 0$.

The first-order condition with respect to $r^L_i$ is

$$\frac{\mu_i \alpha_i}{1 + r^L_i} = \frac{\gamma_i \alpha_i}{\sum_j (1 + r^L_j) \gamma_j \alpha_j}$$

which implies that the optimal labor subsidies follow

$$1 + r^L_i \propto \frac{\mu_i}{\gamma_i}.$$
1.8 Appendix: Robustness of Theory

1.8.1 Alternative Specification of Financial Frictions: Credit Delivery With Linear Monitoring Cost

In this subsection of the appendix, I outline a version of the model with financial frictions taking the form of a linear monitoring cost, which is reminiscent of the implicit wedge formulation in Hsieh and Klenow (2009) and Jones (2013). I abstract away from firm-heterogeneity and entry for expositional clarity, but these modeling elements can be added with notational changes that are conceptually simple. Indeed, it can be shown that this formulation of the model is isomorphic to the model in the main text.

There is a representative consumer who supplies labor inelastically and consumes a unique final good produced competitively according to a CRTS production function $F(\{Y_i\})$. There are $S$ intermediate sectors, each is occupied with a representative producer with CRTS production technology

$$Q_i = h_i F_i (L_i, \{M_{ij}\}).$$

Each intermediate producer faces a working capital constraint

$$\sum_{j \in K_i} p_j M_{ij} \leq W_i$$

where $K_i \subseteq \{1, \cdots, S\}$ is the set of constrained inputs and $W_i$ is an endogenous amount of working capital that is available to the producer in sector $i$.

We introduce a representative financial institution (lender) to the model, who extends working capital to producers and incur a linear cost $\chi_i$ in terms of the final good for every unit of working capital that is given to producer $i$. The lender behaves competitively and makes zero profit, charging a flat interest rate of $\chi_i$ to producer $i$. Producer $i$ therefore solves

$$\max_{W_i, L_i, M_{ij}} \ p_i F_i (L_i, \{M_{ij}\}) - w L_i - \sum_{j} p_j M_{ij} - \chi_i W_i$$

s.t. \hspace{1cm} \sum_{j \in K_i} p_j M_{ij} \leq W_i.

The equilibrium allocation in sector $i$ follows

$$w L_i = \alpha_i p_i Q_i$$

$$p_j M_{ij} = \begin{cases} \sigma_{ij} p_i Q_i & \text{if } j \notin K_i \\ \frac{1}{1 + \chi_i} \sigma_{ij} p_i Q_i & \text{if } j \in K_i \end{cases}$$
\[ \chi_i W_i = \frac{\chi_i}{1 + \chi_i} \sum_{j \in K_i} \sigma_{ij} p_i Q_i \]

Like in our main text, let \( Y \) be the equilibrium net aggregate output and let the expenditure share matrix be \( \Omega = [\omega_{ij}] \) and the elasticity matrix be \( \Sigma = [\sigma_{ij}] \). The influence \( \left( \frac{d \ln Y}{d \ln \chi_i} \right) \) and sales vectors \( \left( \frac{d \ln Q_i}{d \ln \chi_i} \right) \) in this economy are respectively

\[
\mu' = \beta' (I - \Sigma)^{-1}, \quad \gamma' = \frac{\beta' (I - \Omega)^{-1}}{\beta' (I - \Omega)^{-1} \alpha}
\]

and the resource constraint in this economy is

\[ Y = F \left( \{Y_i\} \right) - \sum_i \chi_i W_i. \]

This economy is indeed constrained inefficient despite the fact that credit is competitively allocated by the representative lender. The intuition is similar to that provided in the main text: when downstream producers are constrained, there is a wedge between the marginal social and private return to sectoral spending, either on production inputs or on working capital, and this wedge is exactly captured by the sales gap. To see this, consider labor subsidies \( \tau_i^L \), under which the sectoral allocation of labor, intermediate inputs, and monitoring cost respectively follow

\[
L_i = \frac{\left(1 + \tau_i^L\right) \alpha_i \gamma_i}{\sum_{j} \left(1 + \tau_j^L\right) \alpha_j \gamma_j} L
\]

\[
M_{ij} = \begin{cases} 
\sigma_{ij} \gamma_j Q_j & \text{for } j \notin K_i \\
\frac{1}{1 + \chi_i} \sigma_{ij} \gamma_j Q_j & \text{for } j \in K_i
\end{cases}
\]

\[
\chi_i W_i = \frac{\chi_i}{1 + \chi_i} \left( \sum_{j \in K_i} \sigma_{ij} \right) \gamma_i Y.
\]

By totally differentiating production function \( F_i (\cdot) \), we get

\[
d \ln Q_i = \frac{d \ln F_i}{d \ln L_i} d \ln L_i + \sum_j \frac{d \ln F_i}{d \ln M_{ij}} d \ln M_{ij}
\]

\[
= \alpha_i \left( d \ln \alpha_i + d \ln \left(1 + \tau_i^L\right) + d \ln \gamma_i - d \ln \left( \sum_j \left(1 + \tau_j^L\right) \gamma_j \alpha_j\right) \right)
\]

\[
+ \sum_j \sigma_{ij} \left( d \ln \sigma_{ij} + d \ln \gamma_i + d \ln Q_j - d \ln \gamma_i \right).
\]

Recognizing that \( \alpha_i d \ln \alpha_i + \sum_j \sigma_{ij} d \ln \sigma_{ij} = 0 \) due to \( F (\cdot) \) being CRTS and stacking the equations
using matrix notations, we have

\[ d \ln Q - d \ln y = (I - \Sigma)^{-1} \left( \alpha \circ \left( d \ln \left( 1 + \tau^L \right) - d \ln \left( \sum_j \left( 1 + \tau^L_j \right) \gamma_j \alpha_j \right) \right) \right) \]

where \( \circ \) represents the Kronecker product. From the fact that \( \frac{\partial Q}{\partial y} = \gamma_i \) and \( \frac{\partial Y}{\partial y} \), we get

\[ d \ln y = \beta' d \ln \beta' + \beta' (d \ln Q - d \ln \gamma) = \mu \circ \alpha \circ d \ln \left( 1 + \tau^L \right) \left( \sum_j \left( 1 + \tau^L_j \right) \gamma_j \alpha_j \right). \]

where we have used the fact that \( F(\cdot) \) is CRTS hence \( \beta' d \ln \beta' = 0 \). From here and applying similar arguments as in the proof for Theorem 1.2, we can derive that

\[ SR_i^L = \left. \frac{dC/d \tau^L_i}{dE/d \tau^L_i} \right|_{\tau^L=0, \text{holding } T \text{ constant}} = \frac{\mu_i}{\gamma_i}. \]

Results that parallel other parts of Theorems 1.1 and 1.2 as well as Corollary 1.2 and Proposition 1.4 follow from analogous derivations.

### 1.8.2 Alternative Specification of Financial Frictions: Input-Specific Requirement for Upfront Payment

Consider a more general form of credit constraints:

\[ \sum_{j \in S} \eta_{ij}(\nu) p_j m_{ij} \leq W_i(\nu) \quad \text{for } \eta_{ij}(\nu) \in [0, 1]. \quad (1.8.1) \]

where \( \eta_{ij}(\nu) \) parametrizes the fractional cost for input \( j \) that must be paid upfront out of working capital \( W_i(\nu) \) by firm \( \nu \) in sector \( i \). The constraint formulation (1.1) in the main text corresponds to a special case in which \( \eta_{ij}(\nu) = 0 \) for all \( j \not\in K \subseteq S \) and \( \eta_{ij}(\nu) = 1 \) otherwise.

Under the more general formulation of constraints in (1.8.1), firm \( \nu \)'s problem becomes

\[ \pi_i(\nu) = \max_{\ell, \{m_{ij}\}_{j=1}^S} p_i q_i(\nu, \ell, \{m_{ij}\}) - \sum_{j=1}^S p_j m_{ij} - w \ell \quad \text{subject to } (1.8.1) \]

The firm's first-order conditions are:

\[ w \ell_i(\nu) = \alpha_i(\nu) p_i q_i(\nu) \]
\[ (1 + \lambda_i(\nu) \eta_{ij}(\nu)) p_j m_{ij}(\nu) = \sigma_{ij}(\nu) p_i q_i(\nu) \]
where $\lambda_i(\nu)$ is the Lagrange multiplier on the credit constraint and also the highest interest rate that firm $\nu$ is willing to pay to obtain additional working capital. Under the more general formulation, we no longer have a firm-specific wedge between elasticity and expenditure share on constrained inputs; instead, the wedges are firm-input specific, with

$$\left(1 + \lambda_i(\nu) \eta_{ij}(\nu)\right) = \frac{\sigma_{ij}(\nu)}{s_{ij}(\nu)}.$$  \hspace{1cm} (1.8.2)

Nevertheless, our main theoretical results survive under (1.8.1). To see this, we follow the proof of Proposition 1.2 and derive the semi-elasticity of firm $\nu$'s profit with respect to input and output prices. To distinguish between the effect of a change in the output price and a change in the cost of input produced by the same industry $i$, we again use $C_i$ to denote output price and $p_i$ to denote the cost of input $i$, noting that $C_i = p_i$ in equilibrium. We have

$$\frac{\partial \pi_i(\nu)}{\partial \ln C_i} = C_i q_i(\nu),$$

$$\frac{\partial \pi_i(\nu)}{\partial \ln p_j} = -\alpha_i(\nu) C_i q_i(\nu).$$

By applying the Implicit Function Theorem on the free-entry condition in sector $i$, we can verify that Proposition 1.2 holds under the more general formulation of credit constraints:

$$\frac{\partial \ln C_i}{\partial \ln p_j} = \sigma_{ij} \equiv E_{\nu} \left[ \sigma_{ij}(\nu) \frac{q_i(\nu)}{E_{\nu}[q_i(\nu)]} \right] , \quad \frac{\partial \ln C_i}{\partial \ln w} = \alpha_i \equiv E_{\nu} \left[ \alpha_i(\nu) \frac{q_i(\nu)}{E_{\nu}[q_i(\nu)]} \right].$$

Theorems 1.1, 1.2, Corollary 1.2, and Propositions 1.4 and 1.5 follow similarly, with the private return to tax instrument $\tau_i^l$, defined as the ratio between private marginal product and marginal cost for a marginal change in $\tau_i^l$ starting from the decentralized equilibrium, still being $PR_i^l = \frac{\sigma_{ij}}{\omega_{ij}}$. Note that due to the presence of firm-input specific wedges as in (1.8.2), the private return to $\tau_i^l$ can no longer be written as a weighted average of firm-specific wedges.

Proposition 1.6 and equation (1.30) hold true as they are independent of the microfoundation for the credit constraints.

1.8.3 Alternative Specification of Financial Frictions: Pledgeable Revenue

We can generalize the constraints (1.8.1) even further and enable firms to obtain additional working capital by pledging a fraction $\delta_i$ of firm revenue:

$$\sum_{j \in S} \eta_{ij} p_j m_{ij} \leq W_i + \delta_i p_i q_i \quad \text{for} \ \eta_{ij}, \delta_i \in [0, 1].$$  \hspace{1cm} (1.8.3)
This constraint formulation in (1.8.3) nests the specification in Bigio and La’O (2016), who specify all production inputs, including labor, are subject to the constraint with \( \eta_{ij} = 1 \) for all \( j \in S \) and \( W_i = 0 \) for all \( i \) (note again that whether labor is constrained or not does not affect our main results: we can restate Theorems 1.1 and 1.2 by introducing a sectoral labor wedge):

\[
w_i + \sum_j p_j m_{ij} \leq \delta_i p_i q_i.
\]

Proposition 1.2, a key property in establishing our main results in Theorems 1.1 and 1.2, holds under constraints (1.8.3) only when the within-sector heterogeneity across firms is removed: all firms from sector \( i \) obtain the same productivity \( z_i \), working capital \( W_i \), revenue pledgebility \( \delta_i \), and input-specific requirement for upfront payment \( \eta_{ij} \). To see this, consider firm’s problem under the more general constraint (1.8.3) without heterogeneity:

\[
\pi_i (C_i, w, \{p_j\}) = \max_{\ell_i, m_{ij}} C_i q_i - w_i - \sum_j p_j m_{ij}
\]

s.t. \( q_i = z_i f (\ell_i, \{m_{ij}\}) \)

\[
\sum_{j \in S} \eta_{ij} p_j m_{ij} \leq W_i + \delta_i p_i q_i
\]

where recall that we use \( C_i \) to denote output price of good \( i \) in order to distinguish it with the input price \( p_i \): the two objects are equal in equilibrium, but the sectoral unit cost \( C_i (\cdot) \) is defined as a function of \( p_i \) and prices of other production inputs. Let \( \lambda_i \) be the Lagrange multiplier on the credit constraint, we have

\[
\frac{\partial \pi_i}{\partial \ln C_i} = C_i q_i (1 + \lambda_i \delta_i)
\]

\[
\frac{\partial \pi_i}{\partial \ln p_j} = -\sigma_{ij} C_i q_i (1 + \lambda_i \delta_i).
\]

From here we can again derive \( \frac{\partial \ln C_i}{\partial \ln p_j} = \sigma_{ij} \) as in Proposition 1.2 by applying Implicit Function Theorem on the free-entry condition for sector \( i \). Theorems 1.1, 1.2, Corollary 1.2, and Propositions 1.4 and 1.5 follow analogously.

On the other hand, in presence of within-sector heterogeneity, the elasticity of unit cost of production with respect to input prices is

\[
\frac{\partial \ln C_i}{\partial \ln p_j} = \int \sigma_{ij} (\nu) \frac{q_i (\nu)}{q_i (\nu) d \Phi_i (\nu)} d \Phi_i (\nu)
\]

\[
= \mathbb{E}_\nu \left[ \sigma_{ij} (\nu) - \frac{q_i (\nu) (1 + \lambda_i (\nu) \delta_i (\nu))}{\mathbb{E}_\nu [q_i (\nu) (1 + \lambda_i (\nu) \delta_i (\nu))]} \right]
\]

and Proposition 1.2 fails to hold except in the knife-edge case in which \( \lambda_i (\nu) \delta_i (\nu) \) is the same.
across all firms within sector $i$.

1.9 Appendix: Policy Instruments that Target Firms rather than Sectors

Propositions 1.4 and 1.5 consider policy instruments that effectively relax credit constraints by cost supplying working capital only to firms that have binding constraints, rather than to all firms, and the results are that the ratio between social and private marginal returns to expenditure on working capital is captured by the sectoral sales gap. Our result on sectoral input subsidies (Corollary 1.2) can analogously be generalized to tax instruments that apply to a subset of firms rather than to all firms. To see this, consider subsidies $\left\{ \tau_i^j (\nu) \right\}_{\nu \in \mathcal{F}_i}$ applied to input $j$ for a subset of firms $\mathcal{F}_i$ in sector $i$. To capture marginal changes to these firm-specific subsidies, we parametrize

$$
\tau_i^j (\nu) \equiv f_i^j (\nu; \tau)
$$

with $f_i^j (\nu; 0) = 0$ for all $\nu$ and we assume $f_i^j (\cdot)$ is differentiable in $\tau$. The private return to a marginal change in $\tau$, defined as the ratio between total marginal product captured by firms and the total marginal cost of expending inputs following a marginal change in $\tau$, is

$$
PR \equiv \frac{\mathbb{E}_{\nu \in \mathcal{F}_i} \left[ \sigma_{ij} (\nu) q_i (\nu) \frac{d f_i^j (\nu; \tau)}{d \tau} \right]}{\mathbb{E}_{\nu \in \mathcal{F}_i} \left[ s_{ij} (\nu) q_i (\nu) \frac{d f_i^j (\nu; \tau)}{d \tau} \right]}.
$$

On the other hand, the social marginal return, which can be found by following the procedure in the proofs for Theorems 1.1 and 1.2, is

$$
SR \equiv \xi_i \times PR.
$$

1.10 Appendix: Shrinkage Estimator in Section 1.3

This section of the appendix describes the shrinkages procedure used to estimate $\mathbb{E}_\nu \left[ \epsilon_i (\nu) \right]$, $\mathbb{E}_\nu \left[ k_i (\nu) \right]$, and $\mathbb{E}_\nu \left[ \varphi_i^K (\nu)^{-1} \epsilon_i (\nu) k_i (\nu) \right]$ in equation (1.37) of section 1.3.2. Recall that equation (1.35) is estimated on the sample of SOEs to obtain parameter estimates $\widehat{\eta}_i$ and residuals $\widehat{\epsilon}_i (\nu)$. I then use the estimates $\widehat{\eta}_i$ to obtain elasticity $\overline{\sigma}_i^K (\nu; \widehat{\eta}_i)$ for private firms and recover residuals $\overline{\epsilon} \mathcal{S}_i (\nu)$, which can be interpreted as the product between the private wedge and ex-post TFP shock according to equation (1.36). As a result, the product between $\overline{\epsilon} \mathcal{S}_i (\nu)$ and $k_i (\nu)$ is an estimate of $\varphi_i^K (\nu)^{-1} \epsilon_i (\nu) k_i (\nu)$.

For each of $x \in \{ \overline{\epsilon} \mathcal{S} \cdot k, k, \overline{\epsilon} \}$ I separately apply an empirical Bayes (Morris 1983) procedure to
estimate $E_{\nu}[x_i(\nu)]$, exploiting the cross-industry information in the sample. Specifically, I assume $x_i(\nu)$ follows a log-normal distribution with industry-specific means $\theta_i$ and standard deviation $\zeta$, and the $\theta_i$'s are drawn from a parent Normal distribution with mean $\mu$ and standard deviation $\tau$:

$$\ln x_i(\nu) | \theta_i \sim N(\theta_i, \eta), \ \theta_i \sim N(\mu, \tau).$$  \hspace{1cm} (1.10.1)

All parameters $\{\theta_i\}_{i=1}^S, \eta, \mu, \tau$ are unknown, and the end goal is to estimate $\theta_i$ and $\eta$ in order to form estimates of $E[x_i(\nu)] = \exp(\theta_i + \frac{\eta^2}{2})$. I form estimates of $\eta$ directly as

$$\hat{\eta} = \frac{1}{\sum N_i - S} \sum_{i} \sum_{\nu} (x_i(\nu) - \bar{x}_i)^2$$

where $N_i$ is the number of firms in sector $i$ and $\bar{x}_i = \frac{\sum x_i(\nu)}{N_i}$ is the sample average of $x_i(\nu)$ in the sector. I apply empirical Bayes shrinkage procedure to estimate the industry-specific means $\theta_i$'s as follows.

Under the hierarchical Normality assumptions (1.10.1),

$$\bar{x}_i \sim N(\mu, \frac{\eta}{N_i} + \tau)$$

Let $\hat{\tau}$ be an estimate of $\tau$ and let $w_i \equiv \frac{1}{N_i + \hat{\tau}}$. We form estimates of $\hat{\mu}$ by weighting the sample mean of each industry:

$$\hat{\mu} = \frac{\sum_i \bar{x}_i w_i}{\sum_i w_i}.$$  \hspace{1cm} (1.10.2)

On the other hand, we can form the estimate $\hat{\tau}$ as a function of $\hat{\mu}$:

$$\hat{\tau} = \frac{\sum_i w_i \left\{ \left( \frac{S}{S-1} \right) (\bar{x}_i - \hat{\mu})^2 - \frac{\hat{\eta}}{N_i} \right\}}{\sum_i w_i}.$$  \hspace{1cm} (1.10.3)

The 2-vector $(\hat{\mu}, \hat{\tau})$ is solved as the fixed point of this pair of functions (1.10.2) and (1.10.3). Estimates of $\theta_i$ is formed by shrinking the sample mean $\bar{x}_i$ towards $\hat{\mu}$,

$$\hat{\theta}_i = \left(1 - \hat{B}_i\right) \bar{x}_i + \hat{B}_i \hat{\mu},$$

with weights $\hat{B}_i$ being the relative precision of the naive sample mean estimator (to the precision of $\hat{\mu}$, the estimator for the mean of the upper distribution), appropriately adjusted for the degrees of freedom:

$$\hat{B}_i = \frac{S - 3 - \frac{\hat{\eta}}{N_i}}{S - 1 - \frac{\hat{\eta}}{N_i} + \hat{\tau}}.$$
Lastly, we form the estimate $\widehat{E}_\nu [x_i (\nu)]$ for $x \in \{ \iota \in s \cdot k, k, \iota \}$ as

$$\widehat{E}_\nu [x_i (\nu)] = \exp \left( \tilde{\theta}_i + \tilde{\eta}/2 \right).$$

### 1.11 Appendix: Robustness of Empirical Results in Section 1.3

#### 1.11.1 Alternative Specifications

**Sales Gap** The sales gap measure used in the analysis of section 1.3 is constructed from the final share $\beta$, the observed Chinese input-output table $\Omega$, and the estimated average sectoral wedge $e_i = 1 / 2$, according to equations (1.38), (1.39), and (1.40). The results reported in the main text are based on two assumptions that go into the construction of these objects from data. First, $\beta$ is constructed using only private and public consumption according to (1.42) and it excludes sectoral net exports from the final demand. Second, $\Sigma$ is constructed from $\Omega$ and $e_i$ in (1.38), and while the input-output table $\Omega$ includes both manufacturing and non-manufacturing sectors, our empirical strategy only recovers the sectoral wedges for the manufacturing sector. In the main text I assume that producers in the non-manufacturing sectors are unconstrained with $e_i = 1$ when constructing $\Sigma$ according to (1.38). Our main results are robust to using alternative ways to construct the sales gap measure.

First, recognizing that net exports is an important component of final demand and that it might have played a role for policy design in China, I define $\beta^{NX}$ to include net exports and construct alternative sales gap measure $\xi^{NX}$. Specifically, I define

$$\beta^{NX}_i \equiv \frac{\text{Private and public consumption} + \text{net export of good } i}{\text{Total priv. and public consumption} + \text{net export of all goods}}$$

and construct the sales vector $\gamma^{NX}$, influence vector $\mu^{NX}$, and sales gap vector $\xi^{NX}$ by replacing $\beta$ with $\beta^{NX}$ in (1.38), (1.39), and (1.40).

Second, to overcome the lack of data on $e_i$ for non-manufacturing sectors, I construct a partial input-output table that records input coefficients only for the manufacturing sectors. Specifically, let $M$ denote the set of manufacturing sectors. I define $\Omega^M \equiv [\omega^M_{ij}]_{i,j \in M}$ as the partial input-output table, with

$$\omega^M_{ij} = \frac{\omega_{ij}}{1 - \sum_{j \notin M} \omega_{ij}},$$

and I re-define labor share as

$$\alpha^M_i = \frac{\alpha_i}{1 - \sum_{j \notin M} \omega_{ij}}.$$
That is, I scale up the expenditure shares on labor and manufacturing inputs so that the sum of variable profits and the total expenditure on these inputs add up to one, excluding the expenditures on non-manufacturing goods. I then proceed to construct the elasticity matrix $\Sigma^M$ based on the partial input-output table, which is in turn used to build the alternative sales gap measure, $\tilde{\zeta}^M$.

Our main results are that private firms in sectors with higher sales gaps tend to receive more external loans and pay lower interest rates, and that the sectoral presence of Chinese SOEs is heavily directed towards sectors with larger sales gaps. These results are reflected in columns (4) and (8) of Tables 1.4 and 1.5 as well as columns (2) and (4) of Table 1.6. Appendix Tables 1.11.1 and 1.11.2 show that these results are robust to using the alternative sales gap measures, by respectively replicating our main specifications using $\xi^{NX}$ and $\tilde{\zeta}^M$ instead of $\tilde{\zeta}$ as the sales gap measure.

Interest Rate The two measures of credit market interventions are constructed based on firm-level total liabilities and interest payments made in 2007. While almost all firms in the sample report a positive amount of liability, only 58% report positive interest payments. The results reported in Tables 1.3 and 1.5 are based on the subsample of firms that report positive total interest payments. In Appendix Table 1.11.4 I replicate these results and show that they are robust to using the entire sample of firms.

SOE Share Results reported in Table 1.6 uses SOE’s share of total wage payments as the left-hand-side variable. Appendix Table 1.11.5 replicates columns (4), (6), and (7) of Table 1.6 using SOE’s share of gross value-added and share of revenue as the left-hand-side variable. The results are robust to using either of the alternative measures of SOE share: there is a positive relationship between SOE presence and sectoral sales gap, and the relationship is driven by variations within the subset of sectors that produce material goods.

Dropping Joint-Venture and Foreign Firms The sample of manufacturing firms used for analysis in the main text includes joint ventures between domestic and foreign entities as well as a small sample of firms for which foreign entities hold controlling stakes. I construct alternative sectoral wedge $\left(\varphi_i^{K,NF}\right)^{-1}$ and sales gap $\tilde{\zeta}_i^{NF}$ by dropping all firms with >10% of equity held by foreign entities, including those from Hong Kong, Macau, and Taiwan. Appendix Table 1.11.6 replicates key specifications in the main text using these measures on the restricted sample of firms.

1.11.2 Robustness of Estimator

The estimation of industry-specific output elasticities $\sigma_i^K(\cdot)$ boils down to performing industry-specific regressions (1.35) on the sample of SOEs, while the estimation of industry-specific wedges
A Specification Test for Credit Constraints  A key assumption in my analysis is that credit constraints take the form in (1.31) and that private firms face credit constraints only on capital inputs but not on labor and intermediate materials for production. If labor and material inputs are also subject to credit constraints, there would be wedges that distorts expenditure shares on these inputs from their output elasticities, and we have to account for these wedges when constructing the input-output elasticity matrix $\hat{\Sigma}$. The following exercise can be a specification test for this assumption.

Under the maintained assumption that SOEs are unconstrained, we can relate SOE’s expenditure shares on labor and material inputs to their respective output elasticities according to equation (1.34). I therefore estimate the analogue of equation (1.35), replacing the variable on the left-hand-side to be log-expenditure share on labor and on material inputs, to separately estimate the output elasticities with respect to these inputs. I then form estimates of the hypothetical “labor wedge” and “material wedge” for the private firms using analogous procedures as in equations (1.36) and (1.37). If private firms are indeed unconstrained on these inputs, the estimated wedges should be equal to one for all industries. Appendix Table 1.11.8 shows the result of this exercise by replicating part of Table 1.1. For both of these inputs, the unweighted average sectoral wedges have mean around one across industries while the weighted average sectoral wedges have mean around 1.05. Relative to the wedges on capital inputs, which average to 1.2 for the unweighted and 1.34 for the weighted, the hypothetical wedges on both labor and material inputs are much closer to one and have smaller dispersions around their respective means.

Parametrizing $\sigma^K_i(\cdot)$ as an Industry-Specific Constant  In the main text, the output elasticity of capital inputs $\sigma^K_i(\cdot)$ is parametrized as a second-order polynomial of the input expenditures according to equation (1.35). In Appendix Table 1.11.9 I replicate my main results (columns 4 and 8 of Tables 1.4 and 1.5 as well as columns 2 and 4 of Table 1.6) by parametrizing $\sigma^K_i(\cdot)$ as an industry-specific constant, which corresponds to production functions with a constant output elasticity of capital inputs.
Table 1.11.1: Replication of main results in section 1.3 using $\tilde{\xi^{NX}}$ as the sales gap measure

<table>
<thead>
<tr>
<th>Outcome variable</th>
<th>Debt-To-Capital ratio</th>
<th>Interest rate</th>
<th>SOE share of industry value-added</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Private (1)</td>
<td>Private (3)</td>
<td>(5)</td>
</tr>
<tr>
<td></td>
<td>SOE (2)</td>
<td>SOE (4)</td>
<td></td>
</tr>
<tr>
<td>$\xi_{i}^{NX}$</td>
<td>0.104***</td>
<td>-2.674***</td>
<td>0.528***</td>
</tr>
<tr>
<td></td>
<td>(0.0292)</td>
<td>(0.973)</td>
<td>(0.162)</td>
</tr>
<tr>
<td>$\tilde{(\tilde{\xi}_{i}^{K})^{-1}}$</td>
<td>0.00428</td>
<td>0.903</td>
<td>-0.0130</td>
</tr>
<tr>
<td></td>
<td>(0.0274)</td>
<td>(0.877)</td>
<td>(0.100)</td>
</tr>
<tr>
<td>$(K/Y)_{i}$</td>
<td>0.0110***</td>
<td>-1.104***</td>
<td>-0.00123</td>
</tr>
<tr>
<td></td>
<td>(0.00305)</td>
<td>(0.178)</td>
<td>(0.00140)</td>
</tr>
<tr>
<td>$CapForm_{i}$</td>
<td>-0.0416**</td>
<td>-0.798</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>(0.0189)</td>
<td>(0.678)</td>
<td>(0.179)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.421***</td>
<td>6.929***</td>
<td>-0.318</td>
</tr>
<tr>
<td></td>
<td>(0.0412)</td>
<td>(1.137)</td>
<td>(0.201)</td>
</tr>
</tbody>
</table>

Notes: The table replicates key results from Tables 1.4, 1.5, and 1.6 of the main text, replacing the sales gap measure $\tilde{\xi}_{i}$ used in the main text with $\tilde{\xi^{NX}}$ as defined in 1.11.1. $\tilde{(\tilde{\xi}_{i}^{K})^{-1}}$ is the sectoral private return to capital inputs as defined in equation (1.37). For columns (1) through (4), $(K/Y)_{i}$ is the firm-level capital intensity, defined as the ratio between capital stock and firm revenue. For columns (5) and (6), $(K/Y)_{i}$ is the average capital intensity of firms in the sector. $CapForm_{i}$ is the fraction of sectoral output that is unused in the accounting year and is to be used at a future time. Columns (1) and (2) drop outlier firms with Debt-to-Capital ratio that is either negative or above the 99th percentile. Columns (3) and (4) drop outlier firms with interest rate that is either negative or above the 99th percentile. All specifications drop outlier firms with capital intensity that is either negative or above the 99th percentile. Standard errors in parentheses are clustered at the industry level for columns (1) through (4).
Table 1.11.2: Replication of main results in section 1.3 using $\xi^M$ as the sales gap measure

<table>
<thead>
<tr>
<th>Outcome variable</th>
<th>Debt-To-Capital ratio</th>
<th>Interest rate</th>
<th>SOE share of industry value-added</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Private (1)</td>
<td>SOE (2)</td>
<td>Private (3)</td>
</tr>
<tr>
<td>$\xi_i^M$</td>
<td>0.0617***</td>
<td>0.0305</td>
<td>-1.656***</td>
</tr>
<tr>
<td></td>
<td>(0.0183)</td>
<td>(0.0258)</td>
<td>(0.604)</td>
</tr>
<tr>
<td>$(\delta K)_i^{-1}$</td>
<td>0.0112</td>
<td>0.0317</td>
<td>0.744</td>
</tr>
<tr>
<td></td>
<td>(0.0252)</td>
<td>(0.0321)</td>
<td>(0.804)</td>
</tr>
<tr>
<td>$(K/Y)_i$</td>
<td>0.0110***</td>
<td>0.000296***</td>
<td>-1.106***</td>
</tr>
<tr>
<td></td>
<td>(0.00303)</td>
<td>(0.000108)</td>
<td>(0.178)</td>
</tr>
<tr>
<td>CapForm$_i$</td>
<td>-0.0308**</td>
<td>0.0396</td>
<td>-1.007</td>
</tr>
<tr>
<td></td>
<td>(0.0216)</td>
<td>(0.0240)</td>
<td>(0.662)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.453***</td>
<td>0.544***</td>
<td>6.173***</td>
</tr>
<tr>
<td></td>
<td>(0.0380)</td>
<td>(0.0576)</td>
<td>(1.000)</td>
</tr>
<tr>
<td>Obs.</td>
<td>279060</td>
<td>14211</td>
<td>162143</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.004</td>
<td>0.003</td>
<td>0.026</td>
</tr>
</tbody>
</table>

Notes: The table replicates key results from Tables 1.4, 1.5, and 1.6 of the main text, replacing the sales gap measure $\xi$ used in the main text with $\xi^M$ as defined in 1.11.1. $(\delta K)_i^{-1}$ is the sectoral private return to capital inputs as defined in equation (1.37). For columns (1) through (4), $(K/Y)_i$ is the firm-level capital intensity, defined as the ratio between capital stock and firm revenue. For columns (5) and (6), $(K/Y)_i$ is the average capital intensity of firms in the sector. CapForm$_i$ is the fraction of sectoral output that is unused in the accounting year and is to be used at a future time. Columns (1) and (2) drop outlier firms with Debt-to-Capital ratio that is either negative or above the 99th percentile. Columns (3) and (4) drop outlier firms with interest rate that is either negative or above the 99th percentile. All specifications drop outlier firms with capital intensity that is either negative or above the 99th percentile. Standard errors in parentheses are clustered at the industry level for columns (1) through (4).
Table 1.11.3: Replication of Tables 1.3 to include the sample of firms that report zero interest payments

<table>
<thead>
<tr>
<th></th>
<th>IntRate_i (ν)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td></td>
</tr>
<tr>
<td>1 (SOE_i (ν))</td>
<td>-1.203***</td>
<td>-1.251***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0520)</td>
<td>(0.0516)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>2.638***</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0111)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sector Fixed Effects</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Obs.</td>
<td>276596</td>
<td>276596</td>
<td></td>
</tr>
<tr>
<td>adj. R^2</td>
<td>0.002</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table replicates columns (3) and (4) of Table 1.3 by including firms that report zero interest payments, which were dropped from the analysis reported in the main text. I drop outlier firms with interest rate that is either negative or above the 99th percentile.
Table 1.11.4: Replication of Tables 1.5 to include the sample of firms that report zero interest payments

<table>
<thead>
<tr>
<th>Outcome Variable: Interest Rate</th>
<th>Sample: Private Firms</th>
<th>Sample: SOEs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$\xi_i$</td>
<td>-1.888***</td>
<td>-1.945***</td>
</tr>
<tr>
<td></td>
<td>(0.637)</td>
<td>(0.635)</td>
</tr>
<tr>
<td>$(2K)^{-1}$</td>
<td>0.745</td>
<td>0.741</td>
</tr>
<tr>
<td></td>
<td>(0.641)</td>
<td>(0.651)</td>
</tr>
<tr>
<td>$(K/Y)_i(\nu)$</td>
<td>-0.596***</td>
<td>0.588***</td>
</tr>
<tr>
<td></td>
<td>(0.114)</td>
<td>(0.113)</td>
</tr>
<tr>
<td>$\text{CapForm}_i$</td>
<td>-0.891*</td>
<td>-1.101***</td>
</tr>
<tr>
<td></td>
<td>(0.459)</td>
<td>(0.222)</td>
</tr>
<tr>
<td>Constant</td>
<td>4.736***</td>
<td>3.795***</td>
</tr>
<tr>
<td></td>
<td>(0.814)</td>
<td>(0.843)</td>
</tr>
<tr>
<td>Obs.</td>
<td>264005</td>
<td>264005</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.004</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Notes: The table replicates Table 1.5 by including firms that report zero interest payments, which were dropped from the analysis reported in the main text. All specifications drop outlier firms with interest rate that is either negative or above the 99th percentile. Columns (3), (4), (7), and (8) also drop outlier firms capital intensity that is either negative or above the 99th percentile. Standard errors in parentheses are clustered at the industry level.
Table 1.11.5: Replication of Table 1.6 using SOE share of industry gross value-added and industry revenue

<table>
<thead>
<tr>
<th>Sample</th>
<th>SOE share of industry gross value-added</th>
<th>SOE share of industry revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All industries</td>
<td>Producers of material goods</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$\hat{\xi}_t$</td>
<td>0.365***</td>
<td>0.623**</td>
</tr>
<tr>
<td></td>
<td>(0.125)</td>
<td>(0.235)</td>
</tr>
<tr>
<td>$(\varphi^K_t)^{1}$</td>
<td>-0.0542</td>
<td>-0.0473</td>
</tr>
<tr>
<td></td>
<td>(0.0886)</td>
<td>(0.0953)</td>
</tr>
<tr>
<td>$(K/Y)_t$</td>
<td>-0.00115</td>
<td>0.170*</td>
</tr>
<tr>
<td></td>
<td>(0.00122)</td>
<td>(0.0958)</td>
</tr>
<tr>
<td>$CapForm_i$</td>
<td>0.951</td>
<td>0.468*</td>
</tr>
<tr>
<td></td>
<td>(6.963)</td>
<td>(0.244)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.139</td>
<td>-0.570*</td>
</tr>
<tr>
<td></td>
<td>(0.172)</td>
<td>(0.301)</td>
</tr>
<tr>
<td>Obs.</td>
<td>66</td>
<td>46</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.091</td>
<td>0.079</td>
</tr>
</tbody>
</table>

Notes: The table replicates selected specifications from Table 1.6 using alternative measures of SOE share. Columns (1) through (3) use SOE's share of sectoral gross value-added as the outcome variable, whereas columns (4) through (6) use SOE's share of sectoral total revenue as the outcome variable.
Table 1.11.6: Replication of main results in section 1.3, dropping joint-ventures and firms with foreign ownership

<table>
<thead>
<tr>
<th>Outcome variable</th>
<th>Debt-To-Capital ratio</th>
<th>Interest rate</th>
<th>SOE share of industry value-added</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Private (1)</td>
<td>SOE (2)</td>
<td>Private (3)</td>
</tr>
<tr>
<td></td>
<td>(5)</td>
<td>(6)</td>
<td></td>
</tr>
<tr>
<td>( \hat{\xi}_i )</td>
<td>0.0902***</td>
<td>0.0300</td>
<td>-2.713***</td>
</tr>
<tr>
<td></td>
<td>(0.0291)</td>
<td>(0.0336)</td>
<td>(0.832)</td>
</tr>
<tr>
<td>( (\hat{\varphi}^K)_i^{-1} )</td>
<td>0.0109</td>
<td>0.0295</td>
<td>0.692</td>
</tr>
<tr>
<td></td>
<td>(0.0315)</td>
<td>(0.0337)</td>
<td>(0.878)</td>
</tr>
<tr>
<td>( (K/Y)_i (\nu) )</td>
<td>0.0210***</td>
<td>0.000282***</td>
<td>-1.218***</td>
</tr>
<tr>
<td></td>
<td>(0.00354)</td>
<td>(0.0000994)</td>
<td>(0.195)</td>
</tr>
<tr>
<td>( CapForm_i )</td>
<td>-0.0324</td>
<td>0.0435</td>
<td>-0.731</td>
</tr>
<tr>
<td></td>
<td>(0.0191)</td>
<td>(0.0243)</td>
<td>(0.636)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.424***</td>
<td>0.561***</td>
<td>7.699***</td>
</tr>
<tr>
<td></td>
<td>(0.0499)</td>
<td>(0.0643)</td>
<td>(1.131)</td>
</tr>
<tr>
<td>Obs.</td>
<td>224541</td>
<td>12714</td>
<td>146924</td>
</tr>
<tr>
<td>adj. R(^2)</td>
<td>0.008</td>
<td>0.002</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.096</td>
</tr>
</tbody>
</table>

Notes: The table replicates key results from Tables 1.4, 1.5, and 1.6 of the main text, dropping firms with > 10% equity held by foreign entities from the analysis. \( \hat{\xi}_i \) is the sales gap measure as in equation (1.39). \( (\hat{\varphi}^K)_i^{-1} \) is the sectoral private return to capital inputs as defined in equation (1.37). For columns (1) through (4), \( (K/Y)_i \) is the firm-level capital intensity, defined as the ratio between capital stock and firm revenue. For columns (5) and (6), \( (K/Y)_i \) is the average capital intensity of firms in the sector. \( CapForm_i \) is the fraction of sectoral output that is unused in the accounting year and is to be used at a future time. Columns (1) and (2) drop outlier firms with Debt-to-Capital ratio that is either negative or above the 99th percentile. Columns (3) and (4) drop outlier firms with interest rate that is either negative or above the 99th percentile. All specifications drop outlier firms with capital intensity that is either negative or above the 99th percentile. Standard errors in parentheses are clustered at the industry level for columns (1) through (4).

Table 1.11.7: Number of firms in each industry by firm type

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>St. Dev</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of SOEs in each industry</td>
<td>212</td>
<td>202</td>
<td>135</td>
<td>25</td>
<td>885</td>
</tr>
<tr>
<td>Number of private firms in each industry</td>
<td>4272</td>
<td>4301</td>
<td>2733</td>
<td>222</td>
<td>20557</td>
</tr>
</tbody>
</table>

Notes: The table provides summary statistics on the number of SOEs and private firms in each industry.
Table 1.11.8: Specification test: hypothetical wedges on labor and material inputs

<table>
<thead>
<tr>
<th></th>
<th>$E \left[ \psi_i^L (\nu)^{-1} \right]$</th>
<th>$E \left[ \phi_i^L (\nu)^{-1} \right]$</th>
<th>$E \left[ \psi_i^X (\nu)^{-1} \right]$</th>
<th>$E \left[ \phi_i^X (\nu)^{-1} \right]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.9982</td>
<td>1.0512</td>
<td>0.9971</td>
<td>1.0524</td>
</tr>
<tr>
<td>St. dev</td>
<td>0.078</td>
<td>0.086</td>
<td>0.003</td>
<td>0.003</td>
</tr>
</tbody>
</table>

Notes: The table provides summary statistics for the hypothetical wedges on labor and material inputs and it serves as a specification test of the assumption that only capital goods are subject to credit constraints: the hypothetical wedges on labor and material inputs should be zero for all firms in all sectors. Columns (1) and (3) respectively correspond to the unweighted sectoral average of firm-level wedges on labor and material inputs. Columns (2) and (4) corresponds to the industry average labor and material wedge respectively weighted by the amount of labor and material inputs used by each firm.
Table 1.11.9: Replication of main results in section 1.3, assuming firm-production functions have constant output elasticity for capital inputs

<table>
<thead>
<tr>
<th>Outcome variable</th>
<th>Debt-To-Capital ratio</th>
<th>Interest rate</th>
<th>SOE share of industry value-added</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Private (1)</td>
<td>SOE (2)</td>
<td>Private (3)</td>
</tr>
<tr>
<td>$\hat{\xi}_i$</td>
<td>0.135***</td>
<td>0.0471</td>
<td>-3.736***</td>
</tr>
<tr>
<td></td>
<td>(0.0314)</td>
<td>(0.0419)</td>
<td>(1.158)</td>
</tr>
<tr>
<td>$(\hat{\alpha}_i^{K})^{-1}$</td>
<td>-0.0335</td>
<td>0.00514</td>
<td>1.357</td>
</tr>
<tr>
<td></td>
<td>(0.0322)</td>
<td>(0.0369)</td>
<td>(1.183)</td>
</tr>
<tr>
<td>$(K/Y)_i(\nu)$</td>
<td>0.0108***</td>
<td>0.000295***</td>
<td>-1.101***</td>
</tr>
<tr>
<td></td>
<td>(0.00299)</td>
<td>(0.000106)</td>
<td>(0.180)</td>
</tr>
<tr>
<td>CapForm$_i$</td>
<td>-0.0356</td>
<td>0.0445</td>
<td>-0.773</td>
</tr>
<tr>
<td></td>
<td>(0.0185)</td>
<td>(0.0230)</td>
<td>(0.608)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.434***</td>
<td>0.564***</td>
<td>7.599***</td>
</tr>
<tr>
<td></td>
<td>(0.0558)</td>
<td>(0.0678)</td>
<td>(2.197)</td>
</tr>
<tr>
<td>Obs.</td>
<td>279060</td>
<td>14211</td>
<td>162143</td>
</tr>
<tr>
<td>adj. $R^2$</td>
<td>0.005</td>
<td>0.002</td>
<td>0.028</td>
</tr>
</tbody>
</table>

Notes: The table replicates key results from Tables 1.4, 1.5, and 1.6 of the main text, parametrizing output elasticity with respect to capital as industry-specific constants. $\hat{\xi}_i$ is the sales gap measure. $(\hat{\alpha}_i^{K})^{-1}$ is the sectoral private return to capital inputs as defined in equation (1.37). For columns (1) through (4), $(K/Y)_i$ is the firm-level capital intensity, defined as the ratio between capital stock and firm revenue. For columns (5) and (6), $(K/Y)_i$ is the average capital intensity of firms in the sector. CapForm$_i$ is the fraction of sectoral output that is unused in the accounting year and is to be used at a future time. Columns (1) and (2) drop outlier firms with Debt-to-Capital ratio that is either negative or above the 99th percentile. Columns (3) and (4) drop outlier firms with interest rate that is either negative or above the 99th percentile. All specifications drop outlier firms with capital intensity that is either negative or above the 99th percentile. Standard errors in parentheses are clustered at the industry level for columns (1) through (4).
Bibliography


113


United Nations Department of Economic and Social Affairs (1999). *Handbook of Input-Output Table Compilation and Analysis*.


Chapter 2

Keeping the Little Guy Down: A Debt Trap for Informal Lending

2.1 Introduction

Microcredit was long celebrated for its promise to lift the developing world out of poverty.\textsuperscript{1} Its proponents argued that, by offering a sustainable source of capital, microcredit would enable small scale entrepreneurs to leverage profitable investment opportunities and begin a path to a more prosperous future. These hopes were bolstered by a number of experimental studies that found that many microfirms in the developing world enjoy extremely high marginal returns to capital (on the order of five to ten percent \textit{per month}; see De Mel, McKenzie, and Woodruff (2008), Fafchamps, McKenzie, and Woodruff (2014), and McKenzie and Woodruff (2008)). However, a variety of recent experimental and non-experimental evidence suggests that the impact of microcredit falls far short of previous expectations: on average, firms that randomly receive microcredit are no more profitable than those that do not (see Bangrjic, Karlan, and Zinman (2015) and Meager (2016) for a summary of the recent experimental evidence). A similar puzzle presents itself when considering other forms of informal finance available to these small firms: why haven’t informal financiers such as money lenders enabled these small scale entrepreneurs to leverage their high return to capital opportunities?

The proximate answer for why microcredit, and more broadly informal finance, has so far failed to empower these entrepreneurs may lie in the various contractual features, other than the transfer of working capital, that are common in informal loans.\textsuperscript{2} Many of these features seem to stymie the

\textsuperscript{2}There may be disagreement about whether microfinance should be classified as a form of informal finance. In this paper, we refer to any form of lending as informal if relatively wealthy borrowers are likely to terminate the relationship

\textsuperscript{1}See e.g. the 2006 Nobel Peace Prize awarded to Muhammad Yunus and the Grameen Bank for the innovation and practice of microcredit.

\textsuperscript{2}This chapter is joint work with Benjamin Roth.

117
borrower's ability to invest her loan into high growth opportunities. One prominent example is that many microfinance institutions (MFIs) require that repayment begin immediately after the initial disbursement of funds and take place in frequent installments. The need to have cash on hand may restrict the borrower's ability to undertake long term investments that serve to grow her business at the expense of short term output. Field, Pande, Papp, and Rigol (2013) describe a field experiment in which this restriction was relaxed for a random set of borrowers. Three years after the study took place, borrowers who received two-month grace periods had roughly 80% more business capital and enjoyed 41% higher profits than their counterparts who received standard contracts.

While money lenders are known to allow flexible repayment schedules, they may utilize other means to deter some forms of investment. For instance money lenders commonly require that borrowers work on their land (such as in tenancy arrangements) or that borrowers must forfeit their own land for the money lender to use for the duration of the loan (see e.g. Sainath (1996)). In the former case the lender ties up the borrower's labor, preventing him from focusing on projects to expand his own productive capacity and in the latter case the lender ties up an asset that the borrower could otherwise put to productive use. A final contractual restriction common to both money lenders and microfinance is the use of guarantors (or joint liability, in the case of microfinance). Banerjee, Besley, and Guinnane (1994) theorized that guarantors might pressure borrowers to eschew profitable but potentially risky investments in favor of safer uses of the loan to ensure its repayment. And Fischer (2013) provides evidence from a lab in the field experiment that such pressures indeed exist.

The question then becomes why is the rigid enforcement of these contractual features commonplace? Often these restrictive features are attributed to ensuring the repayment of loans, since formal recourse is unavailable to many informal lenders. That they also restrict investment is largely seen as an unintended consequence. However, with the exception of immediate and frequent repayments, there is little conclusive evidence that these contractual provisions actually serve to reduce default. And in Section 1.6 we suggest that even in the case of immediate and frequent repayments the story may not be clear.

We address both questions raised above by invoking an old explanation: informal lenders may benefit from keeping their borrowers in a debt trap, discouraging them from taking profitable investments to ensure they will continue to borrow for as long as possible (see e.g. Bhaduri (1973), and Bhaduri (1977)). However Braverman and Srinivasan (1981) and Braverman and Stiglitz (1982) argue profit maximizing lenders would not discourage such investments on the grounds that they should allow the efficient level of investment so long as they can extract the surplus. In this paper we argue that relatively wealthy borrowers may leave the informal lending relationship in favor of more attractive sources of financing and if the lender has the capacity to influence the borrower's project selection. We argue below that both of these criteria are satisfied by microfinance.

3 Gine and Karlan (2014) and Attanasio, Augsburg, De Haas, Fitzsimons, and Harrungart (2015) each provide experimental evidence that joint liability does not affect the likelihood of default. In contrast, using non-experimental variation, Carpena, Cole, Shapiro, and Zia (2013) finds that joint liability loans are more likely to be repaid.
when they become eligible for cheaper, more formal sources of credit, providing a natural source of non-transferable utility. For money lenders it may be that their relatively wealthy borrowers become eligible for microloans, and for MFIs it may be that their borrowers become eligible for collateralized bank loans. In either case, relatively poor borrowers may not be able to commit to share the benefits of “formal sector” lending (should they ever reach it) with their informal lender, and thus informal lenders may not be able to extract the surplus of investment. This reopens the possibility that a profit maximizing lender would deliberately impose a debt trap.4

Specifically, our model rests on three critical assumptions which are characteristic of much of the informal lending sector. First, borrowers who save and become sufficiently wealthy cease interaction with their informal lenders and enter the formal sector. While borrowers who engage in the formal lending sector enjoy its benefits, their informal lenders may regret losing customers. Second, borrowers and their informal lenders bargain not only over the division of surplus (i.e. the interest rate of the loan) but also over contractual restrictions which govern the ease with which the borrower can invest her loan to grow her business. Finally, neither the borrower nor the lender can commit to long term contracts. The borrower cannot commit to share the benefits of the formal sector (should she ever reach it) with her informal lender, and the lender cannot commit to provide favorable financing to the borrower in the future in exchange for the borrower’s cooperation in the short term.

For a stylized example, consider a fruit vendor. At low levels of wealth she operates a mobile cart. Each year she receives an endowment of cash and decides how to allocate it between two projects: a working capital project (e.g. buying fruits to sell throughout the week) and a fixed capital project (e.g. buying wood to expand from a mobile cart to a permanent stall from which to sell her fruits). If she invests in working capital, she begins to realize its returns immediately and consumes the proceeds from investment, as for the duration of the period she may not again have the minimum investment required to undertake the fixed capital project. In contrast, if she invests in the fixed capital project, she will forgo consumption for a period but have expanded her capacity for production in future periods. After several more business expansions she will gain access to the formal lending sector, and reap the corresponding benefits.

Each year she is also offered a loan from an MFI (her informal lender). Not only does the loan contract stipulate an initial cash transfer and an interest rate, but it also specifies whether she will be subjected to a variety of measures that make it difficult for her to undertake fixed capital investment and grow her business. For concreteness, assume it specifies whether the borrower must begin repayment immediately, or whether she can wait until the end of the period and repay in one lump installment. By controlling this additional feature, the lender may guide the borrower’s project selection. If the borrower can wait until the end of the year to repay her loan, she may be able to choose her project flexibly. In contrast, if the borrower must have cash on hand to repay

---

4Ray (1998) discusses another explanation for debt traps: borrowers who become sufficiently wealthy may default on their loans, causing lenders to discourage investments that help them grow their business.
her initial installments, she may need to invest in her working capital project at the beginning of the period. And if she has trouble saving cash and unutilized assets (for instance, because she feels pressured to share unutilized assets with family members) then by the time she has earned enough to be able to cover her initial installments she may no longer have the necessary cash to undertake her fixed capital project. That is, the moment at which the lender makes the initial cash transfer is special; this is the moment when the borrower has enough cash to undertake her fixed investment, and if she cannot undertake it immediately she may not be able to for the remainder of the period. Of course, accepting the contract is voluntary, so if the borrower does not find the initial cash transfer and interest rate sufficiently attractive to offset the additional contractual restrictions, she can reject it and allocate her own, smaller endowment flexibly among her projects.

We show that subject to a plausible contracting friction, an asymmetry arises between contracts that restrict the borrower’s ability to grow her business and those that do not. If the lender is unable to set an interest rate which leaves the borrower with exactly the level of output she would have had in his absence, she will retain more utility from unrestricted contracts in equilibrium. This asymmetry arises because she cannot commit to share the proceeds of business growth, and therefore values the investment of her residual income into fixed capital more highly than its investment into working capital. If this asymmetry is sufficiently large the borrower may get stuck in a debt trap; despite large welfare gains from growth, the lender imposes repeated contractual restrictions on the borrower and she remains in poverty (and borrowing from her informal lender) forever. We show such a debt trap occurs if and only if the additional surplus the borrower gains from unrestricted contracts exceeds the additional social welfare generated from business growth.

Beyond establishing that firms offered access to credit often fail to reach their efficient size, the theory organizes a number of other well established empirical facts about microcredit. In our equilibrium, sufficiently wealthy borrowers always receive unrestricted contracts. These are borrowers for whom business expansion is especially valuable due to their proximity to the formal sector, and thus a lender seeking to restrict their investment would need to compensate them with prohibitively low interest rates. This is consistent with many experimental estimates of the impact of microcredit which find that, while the marginal return to microcredit for the average firm is indistinguishable from zero, relatively richer business owners do exhibit high marginal returns to microcredit (see Angelucci, Karlan, and Zinman (2015), Augsburg, De Haas, Harmgart, and Meghir (2015), Banerjee, Duflo, Glennerster, and Kinnan (2015), and Crepon, Devoto, Duflo, and Pariente (2015)).

A further empirical regularity noted in the above experiments is that demand for microcredit contracts is substantially lower than previously expected. This too emerges as a prediction of our theory. Because the lender transfers less surplus to borrowers via restrictive contracts than unrestricted ones, borrowers who receive restrictive contracts may be nearer their indifference condition and their demand for credit may be low.
The model also sheds light on a number of nuanced comparative statics. Improving the attractiveness of the formal sector improves welfare of relatively richer borrowers because they anticipate eventually entering it. On the other hand, this improvement in the formal sector may harm the welfare of poorer borrowers and cause them to be trapped at even lower levels of wealth. Intuitively this is because of a "trickle down" effect whereby lenders anticipate that richer borrowers become more demanding, and restrict the investment of poor borrowers to ameliorate their increased bargaining power. This is especially striking given that fixing any lender behavior, an improvement in the formal sector unambiguously increases the borrower’s welfare. It is because of the lender’s endogenous response that this improvement harms the borrower.

Our comparative static on the borrower’s patience offers a counterpoint to a standard intuition that poverty traps are driven by impatience. In our model, the effect of increasing the borrower’s patience is ambiguous. Increasing the borrower’s patience increases her value of investment, and thus relatively richer borrowers who anticipate eventually entering the formal sector are made better off. However, similar to the comparative static on the attractiveness of the formal sector, increasing the borrower’s patience can make poor borrowers worse off. Anticipating that rich borrowers have improved bargaining power, the lender may restrict the investment of poor borrowers, tightening the debt trap.

Finally we provide two novel empirical facts in support of our theory. The first observation supports one of our key modeling assumptions. Relatively richer borrowers of a large Indian MFI terminate the informal borrowing relationship with higher frequency than their poorer counterparts. This creates potential for the MFI to desire to restrict the business growth of its borrowers to ensure a continued relationship.

Our second observation supports one of the model’s key testable predictions. We provide evidence of our comparative static on competition. During the 2002 Million Baht Program, the Thai government established village funds endowed with one million Baht in each of many villages. Importantly the size of these lending institutions was constant even across villages of varying population size, inducing plausibly exogenous variation in the per capita credit shock. Kaboski and Townsend (2012) leverage this variation to estimate the return to credit for the customers of these village funds. In contrast, we study the variation in contracts offered by money lenders, and argue that Thai money lenders resemble the informal lenders in our model. In particular we show that there is a steep decline in the likelihood a villager borrows from a money lender as a function of his household’s income. Thus we treat the Million Baht Program as exogenous variation in the level of competition faced by these informal money lenders.

We find that in villages with larger per capita credit shocks there is a decline in the incidence of restrictive contracts offered by money lenders, and argue that given the contractual patterns observed in these villages this is the unambiguous prediction of our theory. In essence borrowers in villages with more attractive village funds have better outside options, and the most efficient way
for informal lenders to transfer surplus to their borrowers is to relax contractual restrictions. Thus lenders who previously offered them restrictive contracts at relatively low interest rates now find it unprofitable to do so, and switch to loans which do not restrict the nature of investment. We invoke a number of placebo tests to argue that other theories are unlikely to explain the observed patterns.

A number of other papers offer explanations for the fact that credit markets operate inefficiently. Classic explanations include adverse selection (see e.g. Stiglitz and Weiss (1981)), moral hazard in project selection (Jensen and Meckling (1976)) and moral hazard in repayment (see e.g. Banerjee and Duflo (2010)). Bizer and DeMarzo (1992) suggest that credit markets may operate inefficiently when borrowers cannot commit to exclusive lending relationships and Green and Liu (2016) apply this logic in a development setting to argue that informal lenders may lend the least to borrowers with the most productive investment opportunities. While each of these theories offers an explanation for why credit may not allow firms to fully realize their growth potential, they struggle to match the other empirical regularities we note. Most notably, each of these theories predicts that firms will be credit constrained, and will demand as much credit as they are offered. In contrast our model offers an explanation for the empirical regularity that the demand for microcredit is low.

Finally, several papers offer theories that yield comparative statics of a similar flavor to our own. Petersen and Rajan (1995) argue that credit markets in which there is a high degree of competition for rich borrowers may feature more constrained lending to poor borrowers, as lenders in high competition environments are less able to reap the rewards of investment in poor borrowers. Jensen and Miller (2015) provide a theoretical model of a farmer choosing a level of education for his child. Highly educated children may opt to migrate to the city rather than assisting their parents with farm work, and therefore as the urban returns to education increase the parent may decrease the level of education he allows his child to reach. In both of these models, the comparative static unambiguously harms one of the parties. In contrast, in our model improving the attractiveness of the formal sector only harms poor borrowers by virtue of helping richer borrowers. We expand on this point in section 1.4.3.

The rest of the paper proceeds as follows. In Section 1.2 we describe the model. Section 1.3 characterizes the equilibrium of our game. Section 1.4 describes comparative statics. Section 1.5 discusses some extensions of the model where we relax some of our stylized assumptions. Section 1.6 documents our novel empirical facts. Section 1.7 concludes. All proofs are relegated to the appendix.

2.2 The Model

Players, Actions, and Timing: We study a dynamic game of complete information and perfectly observable actions. There are two players, a borrower (she) and a lender (he). Each
period lasts length $dt$ and players discount the future at rate $\rho$. For analytical convenience we study the continuous time limit as $dt$ converges to 0. The borrower’s business is indexed with a state variable $w \in \{1, \ldots, n + 1\}$ referred to as her business size.

At the beginning of each period the borrower has an endowment $E_w$, which may be augmented by a loan from her lender. She can invest her endowment into two projects: a working capital project $C$ and a fixed capital project $I$. The working capital project $C$ produces consumption goods which she uses to repay her lender and to eat, and the fixed capital project $I$ governs the rate at which her business size increases. The allocation of her endowment between these two projects may be influenced by contractual restrictions imposed by the lender. We defer detailed explanation of financial contracts, and transition between states to the discussion of timing below, after which we map the modeling assumptions to our earlier anecdote about a fruit vendor.

The timing within each period is as follows:

a) The lender offers a contract $\bar{c} = (R, a) \in C \equiv \mathbb{R}^+ \times \{I, C\}$, where $R$ represents the (contractable) repayment from the borrower to the lender, and $a$ represents the contractual restrictiveness.

b) The borrower chooses to accept or reject the contract. Formally she chooses a decision $d \in \{\text{Accept, Reject}\}$.

i. If she rejects the contract:
   
   i. She receives an endowment $E_w > 0$ to flexibly allocate between two projects.
   
   ii. She chooses an amount $c \leq E_w$ to invest into her working capital project $C$.

   A. We assume this project has linear return: $C(w, c) = q_wc$, for some $q_w > 1$ and the borrower consumes this output.

   iii. She chooses an amount $i = E_w - c$ to invest into her fixed capital project $I(w, i)$, the output of which is specified below.

   iv. The lender receives a flow payoff of 0

ii. If she accepts the contract

i. The lender transfers $T_w > 0$ working capital to the borrower, making her endowment $E_w + T_w$.

ii. If $a = C$, the borrower must invest everything in the working capital project. That is, she invests $i = 0$ in the fixed capital project and $c = E_w + T_w$ in the working capital project.

\[\text{Note, we assume that } T_w \text{ is fixed, and therefore do not study the lender's decision of loan size in this paper.}\]
iii. If $a = I$, the borrower must invest $i = E_w + T_w - \frac{R}{q_w}$ in the fixed capital project and $c = \frac{R}{q_w}$ in the working capital project.

iv. The borrower repays $R$ to the lender who receives a flow payoff of $R - T_w$.

c) If the borrower invests $i$ into her fixed capital project $I$, her business size moves from state $w$ to $w + 1$ according to a Poisson process with arrival rate $\frac{1}{\phi_w} dt$ with $\phi_w > 0$ and remains constant otherwise.\(^6\)

d) If the game ever reaches state $n + 1$ both players cease acting.

i. The borrower receives a continuation payoff $U \equiv \frac{u}{\rho}$

ii. The lender receives a continuation payoff $0$

e) Else the period concludes and after discounting the next one begins.

The timing above can be understood through the lens of the example in our introduction. The borrower is a fruit vendor, and at state $w$ she operates a mobile cart. At the beginning of the year she has a cash endowment $E_w$. If she rejects the lender's contract then she flexibly allocates her endowment between her two projects: a working capital project $C$ which can be understood as purchasing fruits to sell during the week, and a fixed capital project $I$ which can be understood as buying raw materials to expand to a market stall from which she may have access to a broader market, improving her productivity. For every unit she invests in the working capital project, she produces $q_w > 1$ units of output. So $q_w$ may be thought of as the markup she enjoys from selling fruits, and $\phi_w$ may be thought of as the cost of fixed investment. The more she invests in fixed capital, the more likely she is to succeed in expanding her productive capacity by moving to state $w + 1$.

If instead she accepts the contract $(R, a)$, the lender transfers $T_w$ working capital to the borrower and her endowment is $E_w + T_w$. The borrower's subsequent investment decision is determined exclusively by the contractual restriction $a \in \{I, C\}$. If the contract specifies that the borrower should invest in the fixed capital project, that is $a = I$, then she invests her entire endowment into fixed capital save for just enough which she invests into working capital to repay her debt. If instead $a = C$ she invests her entire endowment into the working capital project. At the end of the period she repays her debt $R$ to the lender, and they begin anew in the next period.

Though the stylized model above does not include an detailed description of the timing of output within a period, the contractual restriction $a = C$ can be understood as the requirement of early and frequent repayments. If the lender demands that the borrower has cash on hand each day to repay a small fraction of her loan, she may not be able to initially invest in the long term, fixed capital project which may not return output for weeks. By the time she has generated enough

\(^6\)This assumption may be generalized to allow for any transition process in which the probability of transition from $w$ to any other state scales linearly with investment.
income through her working capital project to ensure she can repay each installment, she may no longer have enough cash on hand to meet the minimum required investment in her fixed capital project, as would be the case if she has trouble saving cash from day to day (for instance because she faces pressure from her family to share underutilized assets). In contrast, a borrower uninhibited by a restrictive repayment plan \( a = I \) may invest freely.\(^7\) Therefore we refer to a contract that specifies \( a = I \) as an unrestrictive contract and a contract that specifies \( a = C \) as a restrictive contract.

The borrower’s business expansion is represented by the discrete state space \( \{1, \ldots, n+1\} \). Each state \( w \) represents a different business size (i.e. \( w = 1 \) may be a mobile cart, \( w = 2 \) a fixed stall, \( w = 3 \) a small store and so on). State \( n + 1 \) is a reduced form representation of the formal sector. The borrower enjoys (unmodeled) benefits of formal loans, and the lender receives a 0 continuation payoff having lost his customer. Notably, because the theorems below hold for any fixed investment cost and any number of states, it is straightforward to extend the model to accommodate a continuous state space. We discuss this further in Section 1.5.

**Parametric Assumptions:**

Arguably our most important parametric assumption is on the range of feasible repayment rates \( R \).

**Assumption 2.1.** We assume that the feasible range of repayment rates satisfies \( R \in [T_w, q_w T_w - h_w] \) with \( h_w > 0 \) for all \( w \).

This assumption guarantees that if the borrower accepts the lender’s contract and sets aside \( R q_w \) of her endowment to invest in her working capital project for repayment, the residual endowment she can invest in either project is at least \( E_w + \frac{h_w}{q_w} \) which necessarily exceeds the endowment \( E_w \) she could have invested on her own. This can be motivated in a number of ways. Most straightforwardly, the borrower might be able to hide \( h_w \) from her lender every period, and thus the repayment rate he sets is bounded above by the residual output resulting from the loan, \( q_w T_w - h_w \). Alternatively one could assume that the borrower can renege on her debt in any period, in which case she must find a new lender at cost \( q_w T_w - h_w \). Then the borrower would never repay a debt in excess of this cost.\(^8\)

The repayment ceiling is critical to many of our results below. Because the borrower cannot commit to share the proceeds from business expansion with her lender, she values investment in

---

\(^7\)For technical convenience we assume that upon receiving an “unrestrictive” contract, the borrower must invest her endowment into the fixed capital project. Because we assume below that the borrower values investment in fixed capital more highly than she values investment in working capital, this assumption could be replaced by allowing her to invest flexibly upon receiving an unrestrictive contract with minimal consequence.

\(^8\)In the equilibrium of our model the borrower may not extract positive rents from the lending relationship. Thus, to take this microfoundation seriously, one can ensure that she always finds it profitable to find a new lender in the event of reneging on the first by assuming she receives an additional positive flow utility from interacting with any lender, that is unaffected by which loan she is offered. This can be motivated by an insurance benefit she receives from knowing her lender, that operates independently from the loans she receives every period.
fixed capital more highly than she values investment in working capital. This in turn implies that
the extra endowment \( E_w + \frac{b_w}{q_w} \) the borrower retains induces an asymmetry between the utility she
derives from restrictive and unrestrictive contracts. When this wedge is sufficiently large, the lender
will impose inefficient contractual restrictions on the borrower, trapping her in poverty.\(^9\)

We next assume that both the borrower and lender are risk neutral.

**Assumption 2.2.** Both the borrower and lender enjoy a linear utility of consumption.

Assumption 2.2 implies that if either player receives a sequence of flow consumptions \( \{u_t\} \), their
lifetime utility is

\[
\int_0^\infty e^{-\rho t} u_t dt
\]

**Histories and Strategies:** A history \( \tilde{h}_t \) is a sequence \( \{\tilde{c}_t, d_t, i_t, w_t\}_{t \leq t} \) of contracts, accept/reject decisions, investment allocations and business states at all periods prior to \( t \). We define \( \tilde{H}_t \) to be the set of histories up to time \( t \).

The lender’s strategy is a sequence of (potentially mixed) contractual offers \( \tilde{c} = \{\tilde{c}(\tilde{h}_t)\}_{\tilde{h}_t \in \tilde{H}_t} \),
where \( \tilde{c}(\tilde{h}_t) \in \Delta(C) \) is the probability weighting of contracts he offers the borrower following
history \( \tilde{h}_t \). The borrower’s strategy is a sequence of accept/reject decisions \( d = \{d(\tilde{h}_t, \tilde{c})\}_{\tilde{h}_t \in \tilde{H}_t, \tilde{c} \in C} \)
and investment decisions in the event of rejection \( i = \{i(\tilde{h}_t, \tilde{c})\}_{\tilde{h}_t \in \tilde{H}_t, \tilde{c} \in c} \). Here \( d(\tilde{h}_t, \tilde{c}) \) denotes
the probability the borrower accepts the contract \( \tilde{c} \) following history \( \tilde{h}_t \), and \( i(\tilde{h}_t, \tilde{c}) \) denotes the
investment allocation the borrower undertakes following history \( \tilde{h}_t \) and rejecting contract \( \tilde{c} \).

**Equilibrium:** Our solution concept is the standard notion of Stationary Markov Perfect Equilibrium
(henceforth equilibrium) which imposes that at every period agents are best responding to
one another and that they only condition their strategies on payoff relevant state variables (in this
case, business size). In particular, neither agent has the ability to commit to a long term contract.

Formally, a strategy profile \( (\tilde{c}, d, i) \) is an equilibrium if

a) \( \tilde{c}(\tilde{h}_t) \) is optimal for the lender at every \( \tilde{h}_t \) given the borrower’s strategy \((d, i)\).

b) \( d(\tilde{h}_t, \tilde{c}) \) and \( i(\tilde{h}_t, \tilde{c}) \) are optimal for the borrower at every \( \tilde{h}_t \) and for every contract \( \tilde{c} \) given
the lender’s strategy \( \tilde{c} \).

c) At any two histories \( \tilde{h}_t \) and \( \tilde{h}'_t \) for which \( w \) is the same, we have \( \tilde{c}(\tilde{h}_t) = \tilde{c}(\tilde{h}'_t) \), \( d(\tilde{h}_t, \tilde{c}) =
\tilde{d}(\tilde{h}'_t, \tilde{c}) \), and \( i(\tilde{h}_t, \tilde{c}) = i(\tilde{h}'_t, \tilde{c}) \).

By studying Stationary Markov Perfect Equilibria, we impose that the lender uses an impersonal
strategy: any borrower with the same business size must be offered the same contract. This may
\(^9\)For technical convenience we also require that the repayment level is bounded below, but that it must be larger
than the principle transfer \( T_w \) is unimportant.
be an especially plausible restriction in the context of large informal lenders such as microfinance institutions whose policy makers may be far removed from the recipients of their loans, rendering overly personalized contract offers infeasible.

2.3 Equilibrium Structure

We now describe the borrower and lender's equilibrium behavior and our main results about the structure of the equilibrium. Section 2.3.1 describes the borrower's autarky problem and sets forth an assumption that guarantees the borrower will eventually reach the formal sector (state \( n + 1 \)) in autarky. Section 2.3.2 describes the key incentives of the borrower and lender necessary to understand the structure of the equilibrium. Section 2.3.3 provides our main results: The equilibrium is unique, and under additional assumptions specified below the probability that the lender offers a restrictive contract is single peaked in the state. Thus, the lender's poorest and richest clients may receive unrestrictive contracts and grow faster than they would have in his absence. But borrowers with intermediate levels of wealth receive restrictive contracts every period and find themselves in a poverty trap. Notably, this poverty trap may exist even if the borrower would have reached the formal sector in autarky and even if the discounted utility from expanding to the formal sector is greater than the total surplus generated from investing the total endowment in working capital in every state. In Section 2.3.4 we argue that several well established empirical facts about microfinance can be contextualized through the lens of this equilibrium.

2.3.1 The Borrower's Autarky Problem

First consider the borrower's autarky problem. That is, the economic environment is as specified in Section 1.2, but the borrower is forced to reject the lender's contract at all times (i.e. she must choose \( d(\tilde{h}_t, \tilde{c}) = 0 \) for all histories \( \tilde{h}_t \) and contracts \( \tilde{c} \)).\(^\text{10}\)

Let \( B_w^{\text{aut}} \) be the borrower's continuation value in autarky in state \( w \). This can be decomposed into a weighted average of her flow payoff in the time interval \([t, t + dt]\), and her expected continuation utility at time \( t + dt \). We have

\[
B_w^{\text{aut}} = \max_i q_w (E_w - i) dt + \left(1 - \rho dt\right) \left(\frac{i}{\phi_w} dt B_{w+1}^{\text{aut}} + \left(1 - \frac{i}{\phi_w} dt\right) B_w^{\text{aut}}\right)
\]

(2.3.1)

Fixing the optimal level of investment \( i \) in state \( w \), rearranging, and ignoring higher order terms we have

\[
B_w^{\text{aut}} = q_w (E_w - i) \rho + \frac{i}{\phi_w} B_{w+1}^{\text{aut}}
\]

(2.3.2)

\(^{10}\)Alternatively, one can imagine that the borrower simply does not have access to a lender.
That is, the borrower's autarky continuation value in state $w$ is a weighted sum of her flow consumption $q_w (E_w - i)$ and her continuation value upon increasing business size, $B_{w+1}^{aut}$. Because equation 2.3.1 is linear in $i$ (and equation 2.3.2 is monotone in $i$), the borrower will choose an extremal level of investment. From here on we will use the notation $\kappa_w \equiv \frac{E_w}{q_w}$, which is the maximum speed the borrower can invest in fixed capital and grow in autarky. We have the following proposition about the borrower's autarky behavior.

**Proposition 2.1.** The borrower invests her entire income in every state iff

$$\frac{q_w E_w}{\rho} \leq \left( \prod_{w' = w}^{n} \alpha_{w'} \right) \frac{u}{\rho} \text{ for all } w,$$

where $\alpha_w \equiv \frac{\kappa_w}{\rho + \kappa_w}$.

**Proof.** See appendix. \qed

The borrower's autarky problem has an attractive structure. If she chooses to invest in fixed capital in state $w$ at every period then her continuation utility in state $w$ is $B_w^{aut} = \alpha_w B_{w+1}^{aut}$. That is, she spends a fraction $(1 - \alpha_w)$ of her expected, discounted lifetime in the current state, and a fraction $\alpha_w$ of her expected, discounted lifetime in all future states $w+1$ and onwards. Likewise in state $w$ she anticipates spending a fraction $\prod_{w' = w}^{n} \alpha_{w'}$ of her expected, discounted lifetime in state $w + m$ and onwards if she invests in fixed capital at every period until reaching state $w + m$.

This property is closely related to the Poisson arrival of jumps. Letting $t$ denote the time of the jump and $\lambda$ be the arrival intensity, $v_1$ be the flow utility the borrower enjoys prior to a jump and $v_2$ the flow utility she enjoys post jump, the borrower's utility is represented by:

$$E_t \left[ \int_0^t v_1 e^{-\rho s} ds + \int_t^\infty v_2 e^{-\rho s} ds \right] = \int_0^\infty \left[ \int_0^t v_1 e^{-\rho s} ds + \int_t^\infty v_2 e^{-\rho s} ds \right] e^{-\lambda t} dt$$

$$= (1 - \alpha) \left( \frac{v_1}{\rho} \right) + \alpha \left( \frac{v_2}{\rho} \right)$$

where $\alpha \equiv \frac{\lambda}{\rho + \lambda}$. Thus the borrower's utility from this process can be represented as the convex combination of her lifetime utility from staying in the initial state forever and her lifetime utility from staying in the post-jump state forever, where the weights on each are a function of the intensity of the arrival process. Having established that the borrower invests in fixed capital in all states in autarky if and only if $\frac{q_w E_w}{\rho} \leq \left( \prod_{w' = w}^{n} \alpha_{w'} \right) \frac{u}{\rho}$ for all $w$, we make the following, stronger assumption and maintain it throughout the subsequent analysis.

**Assumption 2.3.** $\frac{q_w E_w + h_w}{\rho} \leq \left( \prod_{w' = w}^{n} \alpha_{w'} \right) \frac{u}{\rho}$ for all $w$.

Assumption 2.3 guarantees that the borrower would prefer to invest her income into fixed capital rather than invest it into working capital for any flow income stream weakly less than $q_w E_w + h_w$. In addition to ruling out an uninteresting case in the analysis, Assumption 2.3 serves to highlight
that the introduction of a lender may cause a poverty trap to emerge despite the autarkic borrower’s eventual entry into the formal sector. That is, business growth among borrowers with access to credit may be lower than growth among their counterparts without access to credit.

### 2.3.2 Relationship Value Functions

We now outline the borrower and lender’s relationship maximization problems and describe their value functions. Let \( B_w \) be the borrower’s equilibrium continuation utility at the beginning of a period in state \( w \), and let \( B_w (\langle R, a \rangle) \) be her equilibrium continuation utility upon receiving the contract \( \langle R, a \rangle \) in state \( w \). Further, define

\[
B_w^{\text{Rej}} = \max_i q_w (E_w - i) dt + (1 - \rho dt) \left( \frac{i}{t} dt B_{w+1} + \left( 1 - \frac{i}{t} dt \right) B_w \right)
\]

to be her equilibrium continuation utility upon rejecting a contract. These functions satisfy

\[
B_w (\langle R, C \rangle) = \max \left\{ (q_w (E_w + T_w) - R) dt + (1 - \rho dt) B_w, B_w^{\text{Rej}} \right\}
\]

and

\[
B_w (\langle R, I \rangle) = \max \left\{ (1 - \rho dt) \left( \frac{E_w + T_w - R}{t} dt B_{w+1} + \left( 1 - \frac{E_w + T_w - R}{t} dt \right) B_w \right), B_w^{\text{Rej}} \right\}
\]

where for both value functions above, the first expression in the brackets corresponds to the borrower’s continuation utility if she accepts the contract \( \langle R, a \rangle \) and the second term corresponds to her continuation utility if she rejects the contract, she is left with her smaller endowment \( E_w \) and chooses her own allocation of investment.

The lender’s value function \( L_w \) in state \( w \) satisfies

\[
L_w = \max_{\langle R, a \rangle} (R - T_w) dt + (1 - \rho dt) \left( L_w + \sum_{a=1}^{n} \frac{E_w + T_w - R}{t} dt (L_{w+1} - L_w) \right)
\]

such that

\[
q_w (E_w + T_w) - R \geq \kappa_w (B_{w+1} - B_w) \quad \text{if} \quad a = C
\]

\[
T_w \leq R \leq q_w T_w - h_w
\]

Note that the lender’s maximization problem and constraints assume the lender never finds it
optimal to offer the borrower a contract she will reject. The lender’s maximization problem also assumes that the borrower accepts any unrestrictive contract. This is the case so long as the borrower would invest her entire income in fixed capital if she were to reject the contract. The borrower accepts a restrictive contract \( (R, C) \) if and only if her value of consuming what the lender offers is weakly higher than that of rejecting the contract and choosing her own allocation of investment, i.e.

\[
q_w (E_w + T_w) - R \geq \kappa_w (B_{w+1} - B_w).
\]

We refer to the above inequality as the borrower’s individual rationality constraint. In equilibrium the lender always offers one of three contracts:

- \( (T_w, I) \) in states \( w \) where the lender’s value function satisfies \( L_{w+1} - L_w \geq \phi_w \) and thus he wants the borrower to expand as quickly as possible. In such states, the lender charges the lowest possible interest rate, \( T_w \).
- \( (q_w T_w - h_w, I) \) in states where the lender prefers unrestrictive contracts but where \( L_{w+1} - L_w < \phi_w \) so that the lender’s preference for expansion is not so strong so as to drive him to offer the borrower a higher than necessary flow payoff. In such states the lender offers the highest possible interest rate, \( q_w T_w - h_w \).
- \( (q_w (E_w + T_w) - \kappa_w (B_{w+1} - B_w), C) \) in states where the lender prefers a restrictive contract, and therefore charges the highest acceptable interest rate.

**Expansion rents**

The lender’s maximization problem illuminates an important force in our model. If the lender offers the borrower a restrictive contract, he optimally offers her the most extractive repayment rate she finds acceptable, denoted by \( R_w \). This repayment rate is determined by the borrower’s indifference condition between accepting the restrictive contract or investing in fixed capital at her autarkic rate. Receiving this contract at every period the borrower’s continuation utility would be

\[
B_w = (q_w (E_w + T_w) - R_w) dt + (1 - \rho dt) B_w = (1 - \rho dt)(\kappa_w dt B_{w+1} + (1 - \kappa_w dt) B_w)
\]

Rearranging and ignoring higher order terms we have

\[
B_w = \frac{\kappa_w}{\rho + \kappa_w} B_{w+1} = \alpha_w B_{w+1}
\]

---

\(^{11}\)It is straightforward to show that in any Stationary Markov perfect equilibrium, either offering the borrower a restrictive contract with the highest acceptable repayment rate or offering her an unrestrictive contract with the highest feasible repayment rate will dominate offering the borrower a contract she would reject.

\(^{12}\)The borrower may invest in working capital in her outside option in equilibrium if she expects a high rate of investment in fixed capital from the lender. This case is dealt with in the appendix, but the analysis does not substantively differ from the above.
That is, if the lender offers the borrower the least generous acceptable restrictive contract the borrower's continuation utility is exactly what it would be if she invested in fixed capital at her autarkic rate.

On the other hand, if the lender offers a maximally extractive unrestrictive contract in every period, the borrower's continuation value will satisfy

\[
B_w = (1 - \rho dt) \left( \frac{E_w + h_w}{\phi_w} dt B_{w+1} + \left( 1 - \frac{E_w + h_w}{\phi_w} dt \right) B_w \right)
\]

Rearranging and ignoring higher order terms we have

\[
B_w = \frac{\gamma_w}{\rho + \gamma_w} B_{w+1} = \beta_w B_{w+1}
\]

where \(\gamma_w \equiv \frac{E_w + h_w}{\phi_w}\) is the rate of expansion the borrower enjoys when she receives a maximally extractive unrestrictive contract, and \(\beta_w \equiv \frac{\gamma_w}{\rho + \gamma_w}\) is the fraction of her discounted lifetime she expects to spend in state \(w + 1\) and onwards if she invests in fixed capital at rate \(\gamma_w\) in state \(w\). Note that if the lender offers the borrower an unrestrictive contract, the borrower's continuation utility is strictly higher than it would be in autarky, because she is allowed to invest strictly more into fixed capital than she would in autarky.

The difference between the borrower's continuation value upon receiving an unrestrictive contract and upon receiving a restrictive one is \((\beta_w - \alpha_w) B_{w+1}\). We refer to this term as the expansion rent in state \(w\). This asymmetry arises because of the ceiling on feasible repayment rates the lender may set. Recall, after transferring \(T_w\) endowment to the borrower, the lender must set a repayment weakly less than \(q_w T_w - h_w\) with \(h_w > 0\). Thus upon accepting a loan and allocating \(R_{q_w}\) to the working capital project for repayment, the borrower necessarily has a larger residual endowment to allocate to either project than she would have had on her own. Because she values investment in fixed capital more highly than she values investment in working capital, she values this extra income more highly when receiving unrestrictive contracts than she does when receiving restrictive ones.

As will be clear in the following sections, this expansion rent is critical for our main results. The lender may prohibit efficient growth by offering the borrower restrictive contracts, and in equilibrium will do so if and only if the expansion rent highlighted above exceeds the change in joint surplus resulting from expansion.

### 2.3.3 Results

We are now in a position to state our first result.

**Proposition 2.2.** An equilibrium exists and is generically unique.
Proof. See Appendix.

The result follows by backward induction on the state. In any state \( w \) the borrower’s accept/reject decision is pinned down by her state \( w \) continuation value \( B_w \) and her state \( w + 1 \) continuation value \( B_{w+1} \). The primary subtlety arises from the fact that the borrower’s welfare in state \( w \) is increasing in the probability the lender offers an unrestrictive contract in \( w \). The more frequently the borrower anticipates unrestrictive contracts in \( w \) the less demanding she will be of restrictive contracts. Formally, we define \( \delta_w (p_w) \equiv p_w \kappa_w + (1 - p_w) \gamma_w \). It is straightforward to show that a borrower who expects a restrictive contract with probability \( p_w \) in state \( w \) will have a continuation utility of \( B_w (p_w) = \frac{\delta_w (p_w)}{\rho + \delta_w (p_w)} B_{w+1} \) which is decreasing in \( p_w \). The lender determines the interest rate associated with restrictive contracts, \( R_w (p_w) \), to solve

\[
\kappa_w (B_{w+1} - B_w (p_w)) = q_w (E_w + T_w) - R_w (p_w)
\]

from which it is immediate that \( R_w (p_w) \) is decreasing in \( p_w \). Thus it may be that when the borrower expects a restrictive contract with certainty the lender strictly prefers to offer an unrestrictive contract, and when the borrower expects an unrestrictive contract with certainty the lender strictly prefers to offer a restrictive contract. In such a case the unique equilibrium involves a strictly interior \( p_w \) and the expansion rent is \( \left( \delta_w - \frac{\kappa_w}{\rho + \delta_w (p_w)} \right) B_{w+1} \).

A second subtlety is due to the possibility that in equilibrium the borrower invests her autarkic endowment in working capital after rejecting the lender’s contract in some state \( w \). Despite Assumption 2.3, the borrower may invest her autarkic endowment in working capital in state \( w \) if she expects to receive sufficiently attractive unrestrictive contracts in state \( w \), which causes \( B_w \) to be near to \( B_{w+1} \) and depresses the value of business expansion. We show that if there is an equilibrium in which the borrower invests her autarkic flow endowment in working capital in state \( w \), this can only be due to the fact that the lender offers her an attractive unrestrictive contract, which is feasible irrespective of the borrower’s autarkic action and therefore occurs across all equilibria. Thus after rejecting the lender’s contract in state \( w \), she invests her autarkic flow endowment in working capital in any equilibrium.

**Equilibrium Contract Structure**

For the remainder of this section and the next we make the following parametric assumptions. We do so for simplicity and ease of exposition but argue in Section 1.5 that their complete relaxation does not change the qualitative lessons to be drawn from the model.

First we assume that the flow working capital output within the relationship is increasing and weakly concave in the state. Let \( y_w \equiv q_w (E_w + T_w) - T_w \).

**Assumption 2.4.** \( y_w > y_{w-1} \) for all \( w \) and \( y_w - y_{w-1} \geq y_{w+1} - y_w \) for all \( w \).
Second we assume that the borrower's autarky endowment \( E_w \), the amount she can hide \( h_w \), and the cost of investment in fixed capital \( \phi_w \) are constant in \( w \).

**Assumption 2.5.** \( E_w = E_{w'} \equiv E \) for all \( w, w' \), \( h_w = h_{w'} \equiv h \) for all \( w, w' \), and \( \phi_w = \phi_{w'} \equiv \phi \) for all \( w, w' \).

Note that Assumption 2.5 allows us to omit the subscripts on \( \kappa, \gamma, \alpha, \) and \( \beta \).

Our next result regards the equilibrium organization of restrictive states and unrestrictive states under the parametric assumptions above.

**Proposition 2.3.** In equilibrium, the probability the lender offers a restrictive contract \( p_w \) is single peaked in \( w \).

**Proof.** See Appendix.

This result implies that in equilibrium the states can be partitioned into three regions of consecutive states: An initial region with only unrestrictive contracts, an intermediate region in which both kinds of contracts are possible, and a final region in which only unrestrictive contracts are offered. In the intermediate region, the probability a restrictive contract is offered is increasing (potentially reaching 1) and then decreasing. This is depicted in the figure below where white states denote unrestrictive states, black states denote restrictive states and grey states denote mixing states.

Borrowers who arrive at a state in which only restrictive contracts are offered never grow beyond it. The next natural question, therefore, is when do such states arise? We have the following result.

**Proposition 2.4.** In equilibrium, the probability the lender offers a restrictive contract \( p_w = 1 \) if and only if

\[
\beta \left( (L_{w+1} + B_{w+1}) - \frac{q_w (E + T_w) - T_w}{\rho} - \phi \right) \leq (\beta - \alpha) B_{w+1}
\]

The left hand side of the above inequality may be loosely understood as the social gain from investing in fixed capital at rate \( E + \frac{h}{q} \) rather than investing everything into working capital. If the borrower invests at rate \( E + \frac{h}{q} \), then she and the lender expect to spend a fraction \( \beta \) of their lifetime in state \( w + 1 \) and onwards. Once in \( w + 1 \) they jointly enjoy continuation values of \( L_{w+1} + B_{w+1} \) but forgo the consumption they could have enjoyed in state \( w \), \( \frac{q_w (E + T_w) - T_w}{\rho} \), and the cost they incur from expansion is \( \beta \phi \). In contrast, the right hand side of the inequality is the borrower's expansion rent: the additional surplus she commands from unrestrictive contracts relative to restrictive ones.
Thus if this expansion rent exceeds the social gain of business expansion, the lender will offer only restrictive contracts, pinning the borrower to the current state.

Note that because of the restrictions on feasible repayment rates, this is not a model of transferrable utility. Thus the left hand side of the above inequality should not literally be interpreted as a change in social welfare. Nevertheless we will sometimes abuse terminology and say that it is socially efficient to invest in fixed capital when the left hand side of the above inequality is positive.

We are now ready to discuss the intuition behind Proposition 2.3. When the borrower is near the formal sector, it is extremely costly to offer her a restrictive contract. For concreteness, consider a borrower in state $n$. A lender who offers this borrower a restrictive contract in every period needs to compensate her with $\alpha\frac{h}{p}$ consumption over the life of the relationship. For $u$ sufficiently high this is prohibitively costly. However, as the borrower becomes poorer it becomes cheaper to offer her a restrictive contract. Consider a borrower who is at state $w$ and who expects unrestrictive contracts in all future states. A lender who offers this borrower a restrictive contract in every period needs to transfer her only $\alpha\beta^{n-w}\frac{u}{p}$ consumption over the lifetime of the relationship. Thus as the borrower becomes poorer it becomes exponentially cheaper to offer her the restrictive contract.

In this intermediate region the expansion rent may become important. As discussed above, when the borrower’s expansion rent exceeds the social gain from business expansion, the lender offers only restrictive contracts, keeping her inefficiently small. Note that this poverty trap is created by the presence of the lender. Assumption 2.3 guarantees that in autarky the borrower would have grown to her efficient size.

Last, as the borrower becomes sufficiently poor her expansion rent $(\beta - \alpha)B_{w+1}$ decreases, as it is tied to her continuation value in the next state. Moreover, because of the concavity of the output from working capital investment, the joint surplus increase from expansion becomes increasingly large as the borrower becomes poorer. Thus sufficiently poor borrowers receive unrestrictive contracts.

We close this section with a discussion of the source of this poverty trap. One crucial feature is that the minimum endowment residual of repayment the borrower enjoys when contracting with the lender, $E + \frac{h}{q_w}$, is strictly larger than the endowment she would have had on her own, $E$. We encode this fact in the following proposition.

**Proposition 2.5.** If $h = 0$, then $p_w = 0$ for any $w$ in which it is socially efficient to invest in fixed capital (i.e. whenever $\beta \left((I_{w+1} + B_{w+1}) - \frac{q_w(E + T_w)}{p} - T_w - \phi\right) > 0$).

When the lender can choose interest rates flexibly enough such that the borrower can be left with exactly the same amount of income that she would have produced alone, he offers unrestrictive contracts in any state in which the social gain from business expansion is positive. When the lender offers a restrictive contract, he gives the borrower just enough consumption to make her indifferent between accepting the contract and rejecting it and investing $E$ in fixed capital. But if the lender
instead offers the borrower a maximally extractive unrestrictive contract, the borrower remains indifferent, because the endowment she can invest into business expansion is exactly what she could have invested on her own. Since the total social surplus increases and the residual surplus accrues to the lender, he prefers unrestrictive contracts.

While the poverty trap disappears when \( h = 0 \) it is important to note that the unique equilibrium still features inefficiently slow business expansion relative to the social optimum. A natural question then, is what contractual flexibility is required to reach the first best level of investment in fixed capital. It is straightforward to verify that equity contracts—contracts that allow the borrower to commit a fraction of her formal sector flow payoff to the lender in exchange for favorable unrestrictive contracts—are sufficient to guarantee first best investment. However this is primarily a theoretical exercise, as the participants of informal financial markets rarely have the capacity to write equity contracts.

2.3.4 Connection to Empirical Evidence

At this point the model already starts to organize much of the empirical evidence on microcredit cited in our introduction. That firms fail to grow from being offered access to microcredit can be understood through the fact that in our model, firms who enter a state where the lender offers restrictive contracts (the black region in the figure above) never leave it, despite the fact that they would have continued to grow in autarky. That is, this is a model in which having access to an informal lender can reduce business growth.

While the microcredit studies listed above find low marginal returns to credit on average, a number of them find considerable heterogeneity in observed returns to credit. In particular they consistently find a long right tail in returns to credit—the largest businesses in areas that randomly received access to microcredit are substantially larger than the largest businesses in areas that did not. Our model sheds light on this heterogeneity to returns as well. Firms at very low and very high business sizes grow faster in the presence of a lender than in his absence (they grow at least at rate \( \gamma \) rather than \( \kappa \)). Whereas firms at intermediate business sizes may not grow at all in the presence of a lender.

In contrast to many other models with credit constrained borrowers, this model offers a novel explanation for the regular finding that demand for microcredit contracts is low. Borrowers in the restrictive region are pushed exactly to their individual rationality constraint—they are indifferent between taking loans and not. While the exact indifference of these borrowers may seem an artifact of the model, the intuition that the lender can push the borrower nearer to her outside option when preventing her from investing in business expansion seems robust. Thus these borrowers may be expected to waver on their decision to accept a loan. In the appendix we discuss an extension to the model in which the lender is incompletely informed about the borrower's outside option.
In equilibrium borrowers in the restrictive region sometimes reject his offer, whereas those in the unrestrictive region never do.

2.4 Comparative Statics

In this section we discuss how the equilibrium changes with respect to a number of comparative statics. Each of them emphasizes an important "trickle down" nature of our model. Namely, changes to the fundamentals of the contracting environment can have nuanced impacts on equilibrium contracts and welfare that vary depending on the borrower's business size. We close this section with comparison to other theories that leverage similar comparative statics.

2.4.1 Comparative Statics on the Borrower’s Continuation Utility $u$ from Entering the Formal Sector.

Increasing the borrower’s continuation value $u$ from entering the formal sector shifts the entire restrictive region leftward. The poverty trap is relaxed for rich borrowers but tightened for poor borrowers. The intuition behind this observation relies on the fact that, when his borrower is rich enough to be in the final unrestrictive region, increasing the attractiveness of the formal sector makes restrictive contracts more expensive for the lender because it increases the borrower’s desire to expand. So the lender shifts towards unrestrictive contracts, relaxing the poverty trap for rich borrowers.

On the other hand, it is precisely this force that causes the lender to tighten the reins on poorer borrowers, increasing the likelihood he offers them restrictive contracts and trapping them at lower levels of wealth. Holding the lender’s strategy fixed, increasing the attractiveness of the formal sector improves the borrower’s bargaining power in all states. However, the richer the borrower is, the more her bargaining power improves because of her proximity to the formal sector. Thus the lender shifts towards restrictive contracts for poorer borrowers, to prevent them from reaching higher levels of wealth where they can exercise their additional bargaining power. This is encoded in the propositions below.

Let $\bar{w} \equiv \arg\max_w \{w : p_w = 1\}$, and $\underline{w} \equiv \arg\min_w \{w : p_w = 1\}$.

**Proposition 2.6.** Increasing the attractiveness of the formal sector relaxes the poverty trap for relatively rich borrowers, but tightens it for poorer borrowers.

That is, $\frac{dp_w}{du} \leq 0$ for $w \geq \bar{w}$ with strict inequality for $0 < p_w < 1$. $\frac{dp_w}{du} \geq 0$ for $w < \underline{w}$ with strict inequality for $0 < p_w < 1$.

*Proof.* See appendix. \(\square\)
The intuition for the above proposition is inextricably linked to the equilibrium effects on welfare, codified in the next proposition.

**Proposition 2.7.** Increasing $u$ weakly decreases the lender's continuation value in all states, and strictly so for $w \leq \tilde{w}$. Increasing $u$ strictly increases the borrower's continuation utility in all states $w \geq w$, but can decrease it in states $w < w$.

That is, $\frac{dH_w}{du} < 0$ for all $w$ with strict inequality for $w \leq \tilde{w}$. $\frac{dR_w}{du} > 0$ for $w \geq w$. For $\rho > \frac{\kappa}{\kappa + \gamma}$, if $p_{w-1} > 0$ then $\frac{dB_w}{du} < 0$ for states $w < w$.

**Proof.** See appendix. □

Though some of the details are cumbersome, the intuition behind these results is instructive. The comparative static for states $w > \tilde{w}$ is most easily understood. Consider the largest state $n$ at which the borrower remains in the informal sector. Increasing $u$ makes it more expensive to offer the borrower a restrictive contract, because she finds investment in fixed capital more valuable. On the other hand, the borrower accepts any unrestrictive contract due to her expansion rent. So the lender finds unrestrictive contracts relatively more attractive and shifts towards them if he previously chose an interior solution.

The borrower's continuation utility increases for two reasons. She benefits from the increased prevalence of unrestrictive contracts, and conditional on entering the formal sector her utility increases. By assumption, the lender at least weakly prefers to offer unrestrictive contracts and since the utility he derives from doing so is unaffected, so is his equilibrium continuation value. This logic extends straightforwardly by backward induction to all states weakly larger than $\tilde{w}$.

The story changes at or prior to $\tilde{w}$. By definition of $\tilde{w}$, the lender offers a restrictive contract with certainty (i.e. $p_{\tilde{w}} = 1$), and therefore transfers $\alpha B_{\tilde{w}+1}$ utility to the borrower over the lifetime of the relationship. Having already established that the borrower's continuation utility in state $\tilde{w} + 1$ increases with $u$, we can now see that the borrower's utility also increases in state $\tilde{w}$. However, now the increase in her utility results from a direct transfer from the lender, so his continuation utility in state $\tilde{w}$ decreases. A similar conclusion is reached for all states $w \in \{w, \tilde{w}\}$ by backward induction.

Finally, consider state $w - 1$, in which, by definition, the lender at least weakly prefers to offer an unrestrictive contract. Further, suppose that the preference is indeed weak, so that $p_{w-1} \in (0, 1)$ (i.e. the lender offers a restrictive contract with positive probability). First, note that since the borrower never grows beyond state $w$, increasing $u$ has no effect on social welfare in state $w - 1$. However, since the borrower's equilibrium continuation utility in state $w$, $B_w$ increases, so does her expansion rent in state $w - 1$. Recall that her state $w - 1$ expansion rent $\left(\beta - \frac{\kappa}{\rho + \delta(p_{w-1})}\right)B_w$ is a fraction of her continuation utility in state $w$. Because the borrower's share of surplus from business expansion increases but the change in social surplus accruing from business expansion does

137
not, the lender shifts towards restrictive contracts, slowing down the borrower’s growth.

Another way to understand this is the increase in the attractiveness of the formal sector trickles down and increases the borrower’s bargaining power in state $w$. Markov Perfection prevents her from committing not to exercise this additional bargaining power and, because state $w - 1$ borrowers are less affected, they become relatively more attractive and the lender shifts towards offering them restrictive contracts.

How this affects the borrower’s state $w - 1$ equilibrium continuation utility $B_{w-1}$ is in general ambiguous. That $B_w$ increases is a force towards increasing $B_{w-1}$. However, the rate at which she grows to state $w$ slows, which is a force towards reducing $B_{w-1}$. For sufficiently impatient borrowers the latter force dominates, as impatience amplifies the difference between slow and fast rates of expansion, and $B_{w-1}$ decreases in the attractiveness of the formal sector. The lender is made unambiguously worse off from the increase in $u$, because he weakly prefers unrestrictive contracts and his state $w$ continuation utility is decreasing in $u$. Thus an increase in the attractiveness of the formal sector causes a Pareto disimprovement. Because the borrower cannot commit to forgo her improved bargaining power in state $w$, the lender traps her in state $w - 1$ to both of their detriments. And the story continues in much the same way for all states prior to $w - 1$.

That increasing $u$ can make the borrower worse off at some business state $w$ is especially striking in light of the following consideration. Fix any (potentially non-equilibrium) lender behavior characterized by $\{p_w\}$, such that in state $w$ the lender offers a restrictive contract with probability $p_w$ and an unrestrictive contract with probability $1 - p_w$. Then increasing $u$ unambiguously makes the borrower strictly better off in all states. The restrictive contract in state $w$ becomes more generous when $B_{w+1}$ improves, and the borrower’s utility from receiving an unrestrictive contract in state $w$ improves when $B_{w+1}$ increases. This logic is codified in the following proposition.

**Proposition 2.8.** Fixing the lender’s behavior characterized by $\{p_w\}$ defined above, increasing $u$ strictly improves the borrower’s continuation utility in all states.

*Proof.* See appendix.

So fixing the lender’s behavior, regardless of what that behavior is, increasing $u$ unambiguously improves the borrower’s continuation utility. It is because of an equilibrium adjustment to the lender’s behavior, namely that he shifts towards restrictive contracts, that impatient borrowers are made worse off at all business states $w < w$.

### 2.4.2 Comparative Statics on the Borrower’s Level of Patience

A standard intuition about poverty traps is that they are driven by impatience. However in this model increasing patience has a very similar effect to increasing the attractiveness of the formal sector, and hence can tighten the poverty trap and make the borrower worse off at some levels of
wealth. Let \( \rho^B \) be the borrower’s level of patience and \( \rho^L \) be the lender’s level of patience (and note that decreasing \( \rho^B \) is equivalent to increasing patience).

**Proposition 2.9.** Increasing the borrower’s patience relaxes the poverty trap for relatively rich borrowers, but may tighten it for poorer borrowers.

That is, \( \frac{d \rho_w}{d \rho^B} > 0 \) for \( w > \bar{w} \) with strict inequality for \( \rho_w > 0 \). For \( w < \bar{w} \), the sign of \( \frac{d \rho_w}{d \rho^B} \) is ambiguous.

**Proof.** See appendix.

For rich borrowers above the highest pure restrictive state \( (w > \bar{w}) \), the comparative static on \( \rho^B \) works in exactly the same way as the comparative static on \( u \). Increasing the borrower’s patience increases how much she values business expansion. This causes her to be more demanding of restrictive contracts, but leaves the lender’s payoff from offering unrestrictive contracts unchanged. Thus, in all such states the lender shifts towards unrestrictive contracts, increasing the rate that these rich borrowers reach the formal sector.

For borrowers in pure restrictive states \( (w \in \{\bar{w}, \ldots, \bar{w}\}) \), the comparative static on the borrower’s patience again works as it did for changes in the attractiveness of the formal sector. The borrower’s continuation utility in state \( w + 1 \), \( B_{w+1} \), increases so the amount of consumption she demands in return for contractual restrictions increases. This increases her welfare at the direct expense of the lender’s.

Finally, consider state \( w - 1 \). Recall the borrower’s expansion rent in this state is \( \left( \beta - \frac{\rho^B + \delta}{\rho_w} \right) B_w \). That her utility in state \( w \), \( B_w \), increases is a force towards increasing her expansion rent. However, as she becomes more patient, the difference she perceives between slow and fast rates of expansion is muted. That is \( \frac{d \left( \beta - \rho^B + \delta \rho_w \right)}{d \rho^B} > 0 \), which is a force towards decreasing the expansion rent. Which of these two forces dominates is in general ambiguous, but we show in the appendix that these forces can resolve in favor of increasing the expansion rent. Thus, in contrast to standard models of poverty traps, increasing the borrower’s patience can make this poverty trap worse.

### 2.4.3 Comparison to Petersen and Rajan (1995) and to Jensen and Miller (2015)

There are a number of principal agent models that feature a comparative static similar to ours. Two such theories are those of Petersen and Rajan (1995) and Jensen and Miller (2015). Petersen and Rajan study a credit market, and argue that if competition for rich borrowers becomes more fierce, lenders may invest less in their current borrowers. Interpreting this theory through the lens of our model, this is akin to reducing the attractiveness of the formal sector for the lender without changing the attractiveness of the formal sector for the borrower. Doing so reduces the lender’s
payoff from unrestrictive contracts and weakly increases the equilibrium probability of restrictive contracts in all states, making the borrower weakly worse off.

Jensen and Miller (2015) study Indian agricultural households in which parents decide a level of education for their children, and then children decide whether to stay at home and work on the farm or migrate to the city, leaving their parents behind. While education has positive returns in both locations, it has higher returns in the city. They show, both theoretically and empirically, that reducing the cost of migrating to the city causes parents to reduce educational investment in their children. Interpreting their insight in the language of this paper, theirs is an exercise in reducing the wealth level required to enter the formal sector in a model where the lender doesn’t need to respect the borrower’s individual rationality constraint. This is eminently sensible for a parent who chooses a child’s level of education, but may be less so for a lender offering a borrower a loan. Like Petersen and Rajan, the result of their comparative static is to make the borrower weakly worse off in all states.

In contrast to both theories, the welfare of rich borrowers in our model is unambiguously improved by improvements in the formal sector. The sources of improvement are twofold: borrowers who reach the formal sector receive higher utility, and borrowers enjoy more frequent unrestrictive contracts because restrictive contracts become relatively more expensive for the lender. This improvement in the rich borrower’s welfare (and her inability to commit not to exercise her improved bargaining power) is absent in both theories highlighted above, yet is critical to our result that poor borrowers can be made worse off. The lender ameliorates this improvement in bargaining power by transitioning from away from unrestrictive contracts, harming the borrower’s equilibrium welfare.

2.5 Extensions and Robustness

In this section we argue that the key intuitions highlighted above survive a number of extensions to the model. We begin by completely relaxing parametric Assumptions 2.4 and 2.5 and arguing that the qualitative lessons are unchanged.

We then extend the model to study direct competition. We show that if the incumbent lender has a sufficiently large lending advantage, the results are unchanged relative to the monopolist case above. If instead the incumbent has no advantage, the borrower necessarily reaches the formal sector in finite time in equilibrium. We further derive a sufficient condition for a monotone comparative static in the incumbent’s lending advantage which provides a testable prediction for the empirical exercise in Section 1.6.

In the appendix we explore several other extensions. First we allow for the lender to be incompletely informed about the borrower’s outside option and show that in equilibrium the borrower sometimes rejects the lender’s offer of a restrictive contract, providing an explanation for the low
demand of microcredit. We then discuss an extension in which we allow the borrower to flexibly allocate a fraction of her income irrespective of contractual restrictions, and show that the lender may still restrict the rate at which the borrower grows relative to autarky.

2.5.1 Arbitrary production functions

In this section we relax Assumptions 2.4 and 2.5 and discuss how it affects our results. In particular we make no assumptions about $q_w$, $E_w$, $T_w$, $\phi_w$ or $h_w$ other than that $h_w > 0$ for all $w$ and Assumption 2.3 above, which guarantees that in autarky the borrower reaches the formal sector in finite time.

Structure of the Equilibrium

First, recall that Proposition 2.2 which states that the equilibrium is unique was shown without Assumptions 2.4 and 2.5 and thus continues to hold. In this section we discuss the structure of the unique equilibrium. A typical equilibrium is depicted below, with each circle representing a state and shaded circles representing states in which restrictive contracts are offered.

![Diagram of equilibrium structure]

Even though in general we cannot say anything about the organization of restrictive and unrestrictive states, we argue that many of the empirical facts discussed in Section 1.4 can still be understood through the equilibrium above. In fact, with the exception of heterogeneity in returns to credit, our explanation of the facts in that discussion only depended on the potential for each type of contract to coexist in a single equilibrium. As such we focus our attention for the remainder of this discussion on the prediction that wealthy borrowers will receive unrestrictive contracts and thus will enjoy high returns to credit.

To do so, we first outline how to transfer the insights in the above model to one with a countably infinite number of states. Given that our results do not depend on the number of states $n$, or the cost of investment $\phi_w$, this is a straightforward task. We define a sequence of games satisfying the above assumptions, each with successively more states.

Let $\Gamma^1$ be an arbitrary game satisfying the assumptions, with $n$ business states.

For $m > 1$, let $\Gamma^m$ be constructed in the following way:

- $\Gamma^m$ has $2^{m-1}n$ business states, and let $q^m_w$, $E^m_w$, $T^m_w$, $\phi^m_w$, and $h^m_w$ be the corresponding parameters for game $\Gamma^m$. 

141
• If \( w \) is even, set \( q_w^m = q_{w/2}^{m-1}, E_w^m = E_{w/2}^{m-1}, T_w^m = T_{w/2}^{m-1} \), and \( h_w^m = h_{w/2}^{m-1} \)

• If \( w \) is odd, set

\[
q_w^m \in \left\{ \min \{ q_{w-1}^m, q_{w+1}^m \}, \max \{ q_{w-1}^m, q_{w+1}^m \} \right\}
\]

\[
E_w^m \in \left\{ \min \{ E_{w-1}^m, E_{w+1}^m \}, \max \{ E_{w-1}^m, E_{w+1}^m \} \right\}
\]

\[
T_w^m \in \left\{ \min \{ T_{w-1}^m, T_{w+1}^m \}, \max \{ T_{w-1}^m, T_{w+1}^m \} \right\}
\]

and

\[
h_w^m \in \left\{ \min \{ h_{w-1}^m, h_{w+1}^m \}, \max \{ h_{w-1}^m, h_{w+1}^m \} \right\}
\]

• If \( w \) is even, set \( \phi_w^m = \phi_{w/2}^{m-1} / 2 \) and if \( w \) is odd, set \( \phi_w^m = \phi_{(w-1)/2}^{m-1} / 2 \)

Thus \( \Gamma^m \) has twice as many states at \( \Gamma^{m-1} \), and even states in \( \Gamma^m \) correspond to states in \( \Gamma^{m-1} \). The parameters in odd states take values intermediate to those in the surrounding states. Because the cost of investment in \( \Gamma^m \) is only half that in \( \Gamma^{m-1} \), a borrower investing in fixed capital at the same rate in either game would reach the formal sector in the same expected time. One way to understand \( \Gamma^m \) relative to \( \Gamma^{m-1} \) is that the borrower and lender appreciate more nuanced differences in the borrower’s business size. Holding investment rate fixed, it takes the same amount of time to get from \( w \) to \( w + 2 \) in \( \Gamma^{m+1} \) as it does to get from \( w \) to \( w + 1 \) in \( \Gamma^m \), but along the way in \( \Gamma^m \) the borrower and lender realize an intermediate production function change. For \( m' > m \), we say \( \Gamma^{m'} \) is descended from \( \Gamma^m \) if there is a sequence of games \( \Gamma^m, \ldots, \Gamma^{m'} \) that can be derived in this manner. We have the following result.

**Proposition 2.10.** For any \( \Gamma^m \), there is an \( \bar{m} \) such that for all \( m' > \bar{m} \), the equilibrium in any \( \Gamma^{m'} \) descended from \( \Gamma^m \) features a \( \bar{w} \) such that for \( w \geq \bar{w} \) the borrower reaches the formal sector in finite time starting from state \( w \) if it is socially efficient to do so.

**Proof.** See appendix. \( \square \)

The above result says that for any game with sufficiently fine discrimination between states, all sufficiently wealthy borrowers receive unrestricted contracts in equilibrium, and thus realize high returns to credit. The intuition is simple. Because entering the formal sector is efficient, the lender is unable to offer a sufficiently wealthy borrower (one who is sufficiently near to the formal sector) a restrictive contract she will accept. As the borrower and lender become arbitrarily discerning of different states, there will eventually be business states where the borrower is indeed sufficiently wealthy.

**Comparative statics**

As before, we can be fairly precise in describing how the equilibrium changes with respect to various fundamentals of the game. In this section we focus on the comparative static with respect
to $u$.

Note that without loss of generality we can identify $m$ disjoint, contiguous sets of states \{w_1, \ldots, w_1\}, \ldots, \{w_m, \ldots\} such that $\bar{w}_m = \max \{w : p_w = 1\}, \bar{w}_m = \max \{w : p_w = 1, p_{w-1} < 1\}$, and in general for $k \geq 1$, $\bar{w}_k = \max \{w < w_{k+1} : p_w = 1\}$ $w_k = \max \{w \leq \bar{w}_k : p_w = 1, p_{w-1} < 1\}$. An arbitrary set \{w_k, \ldots, \bar{w}_k\} is a contiguous set of states where restrictive contracts are offered with probability 1, and each pure restrictive state is contained in one of these sets.

We consider an impatient borrower and establish the following result.

**Proposition 2.11.** For impatient borrowers, the regions of contiguous restrictive states merge together as the formal sector becomes more attractive.

That is, for $\rho > \max_w \frac{\kappa_w u_m}{\kappa_w + \gamma_w}$, $\frac{dp_w}{du} < 0$ for $w \in \{\bar{w}_m + 1, n\}$, $\frac{dp_w}{du} > 0$ for $w \in \{\bar{w}_{m-1} + 1, u_{\bar{w}_m} - 1\}$, $\frac{dp_w}{du} < 0$ for $w \in \{w_{m-2} + 1, w_{\bar{w}_{m-1}} - 1\}$ and so on.

**Proof.** See appendix.

Proposition 2.11 states that the highest region of pure restrictive states moves leftward, the second highest region moves rightward and so on. This is depicted in the following figure.

The intuition is as follows. For $(w_m, \bar{w}_m)$, the analysis exactly follows that of Section 1.4.1, and hence it shifts leftward as $u$ increases. But recall that for the impatient borrowers to the left of a restrictive state, the leftward shift lowers their utility. This is akin to lowering the utility of entering the formal sector, and hence for the next set of restrictive states $(w_{m-1}, \bar{w}_{m-1})$ the analysis reverses and $\bar{w}_{m-1}$ and $w_{m-1}$ shift rightward. The rest follows by backward induction.

### 2.5.2 Direct Competition

Throughout the above analysis we have assumed the lender is a monopolist. In this section we introduce the possibility of a second lender who can make offers to the borrower. We label one
lender the incumbent and one lender the entrant. The timing and technologies are the same as above, however now each period both lenders offer the borrower a contract. If the borrower accepts the incumbent’s contract everything proceeds as above. If the borrower accepts the entrant’s contract, everything proceeds as above except that the borrower incurs a non-pecuniary penalty of $\psi dt > 0$. This penalty can be understood as a lending disadvantage the entrant suffers relative to the incumbent, perhaps because the incumbent is better equipped to screen or monitor its borrowers and thus borrowers interacting with the entrant undergo more costly screening processes.

We are now ready to state our first result.

**Proposition 2.12.** There exists a unique equilibrium in which the borrower accepts the incumbent’s loan offer in all periods.

*Proof.* See Appendix.

The proof of the above result proceeds much in the same way as the proof of Proposition 2.2, except that rather than the borrower’s outside option being that she can flexibly invest her endowment $E_w$, her outside option may now be to receive an attractive loan from the entrant and incur the non-pecuniary cost of $\psi dt$. Specifically, because the entrant never expects the borrower to accept his loan in the future, he is willing to offer the borrower any loan she would accept in the current period. Thus, in effect the borrower’s outside option is whichever she prefers between flexibly allocating $E-w$, and flexibly allocating $E_w + T$ but incurring the non-pecuniary cost $\psi dt$.

Our next proposition aims to demonstrate that the intuitions derived under the monopolist case survive to the case with two lenders so long as $\psi$ is large enough. In contrast, as $\psi$ becomes sufficiently small, the equilibrium approaches the first best level of business expansion.

**Proposition 2.13.** There exists a $\bar{\psi} > 0$ such that for $\psi > \bar{\psi}$, the equilibrium probability $p_w$ that the incumbent lender offers the borrower a restrictive contract in state $w$ is the same as in the model with one lender.

So long as it is efficient to invest in business growth, there exists a $\bar{\psi} > 0$ such that for $\psi < \bar{\psi}$, the equilibrium probability $p_w$ that the incumbent lender offers the borrower a restrictive contract in state $w$ is 0. And as $\psi \to 0$, the equilibrium repayment rate $R$ charged by the incumbent lender converges to $T_w$ in each state $w$.

For a sufficiently strong lending advantage, the incumbent lender behaves as a monopolist. While this is intuitive, it may nevertheless be important to highlight that de facto monopoly power need not arise from being the only lender available. It may also arise from having low screening and monitoring costs relative to one’s competitors. Proposition 2.13 also asserts that as the incumbent’s lending advantage vanishes, he offers increasingly generous unrestricted contracts and the borrower’s rate of business expansion approaches first best. This serves to highlight that in the limit of perfect competition, the inefficiencies highlighted in the model vanish.
Finally, we discuss comparative statics with respect to the incumbent's lending advantage. While the model is suitable to study the effect of a localized increase in competition across many scenarios, we highlight only one, which motivates the empirical exercise to follow.

**Proposition 2.14.** If the equilibrium is characterized by a \( \tilde{w} \) such that \( p_w = 1 \) for \( w \leq \tilde{w} \), and \( p_w \in (0, 1) \) for \( w > \tilde{w} \), then \( \frac{\partial p_w}{\partial w} \geq 0 \) for \( w > \tilde{w} \).

Proposition 2.14 states that if the equilibrium structure is such that there are a set of states in which the incumbent lender offers restrictive contracts with probability 1 and mixes between the two types of contracts with strictly interior probability at all higher states, then the comparative static with respect to the level of competition is unambiguous. As the incumbent's lending advantage diminishes, the probability he offers restrictive contracts weakly declines in all states, and strictly so in states in which he had previously been offering a restrictive contract with positive probability. As competition increases, the incumbent is forced to transfer more surplus to his borrower, and the most efficient way to do so is to relax the contractual restrictions he imposes. In section 1.6.2, we provide empirical evidence for this testable prediction.\(^\text{13}\)

### 2.6 Novel Empirical Support

In this section we present two novel empirical observations in support of our theory. The first observation supports one of our key modeling assumptions; relatively richer borrowers are more likely to cease interaction with an MFI than are their poorer counterparts. Second we provide empirical evidence for a comparative static of our model. Leveraging exogenous variation in the level of competition facing money lenders in Thai villages, we show that competition causes a relaxation of the contractual restrictions they impose. We further argue that this is the unambiguous prediction of the model given the equilibrium contract structure we observe in the data. As discussed above, this comparative static prediction arises from the fact that money lenders in competitive markets face pressure to transfer additional surplus to their borrowers, and the primary method by which they do so is to ease contractual restrictions.

### 2.6.1 Rich Borrowers Cease Interaction With an MFI

The data from this exercise are from Field et al. (2013). Their paper reports on an experiment conducted in partnership with Village Financial Services, a large Indian MFI. As discussed in the introduction, the authors randomly relaxed the contractual requirement that borrowers begin repayment immediately after loan disbursement; a random subset of borrowers received a two month grace period during which they did not need to meet any repayment obligation, and after which

---

\(^{13}\) The comparative static in the empirical exercise may be more closely modeled by increasing the amount of money the entrant has to lend to the borrower. This would have exactly the same impact on the lending environment as reducing \( \psi \).
repayment took place in standard installments. The authors found that three years after the initial loan disbursement, borrowers in the treatment group reported weekly profits between Rs. 450 and Rs. 900 more than those in the control group. However they also report that borrowers in the treatment group default significantly more on average. While the probability that any amount of money remains in default at the end of the loan cycle more than quadruples (from about 2% in the control group to 9% in the treatment group), the increase in average amount in default is much more modest. On average borrowers in the treatment group default on an additional Rs. 150. Put another way, for a one time additional cost of Rs. 150, the grace period increased weekly business profits by Rs. 450 - Rs. 900. Why then, have MFIs (including Village Financial Services) maintained their strict repayment schedules that begin immediately after loan disbursal?

We propose that the answer lies in the pattern of default as a function of business size. The red line in Figure 1-1 below plots the expected amount in default a year after loan origination as a function of business profits three years after loan origination. As may be expected, there is a clear decreasing relationship between a borrower's level of wealth and the amount of money she is expected to have in default.

However the blue line in Figure 1-1 plots the likelihood a borrower defaulted on any amount of money as a function of business profits three years after loan origination. In this case there is a U-shaped relationship between profits and default; those with relatively low and relatively high business profits are the most likely to default. Table 1.1 presents this pattern in regression form. Specifically we regress

\[ Default_i = \alpha + \beta_1 \ln\text{profi}_i + \beta_2 \ln\text{profsq}_i + \gamma X_i + \epsilon_i \]

where \( Default_i \) is an indicator taking a value of 1 if borrower \( i \) has not completed repayment of her loan a year after origination (and substantially after the final tranche was due), \( \ln\text{profi} \) is the log of borrower \( i \)'s business profits three years after loan origination, \( \ln\text{profsq} \) is the square of her profits three years after origination and \( X_i \) is the vector of controls used in Field et al. (2013). As can be seen, \( \beta_1 \) is negative, \( \beta_2 \) is positive, and both are statistically significant once the vector of controls is included.

Taken together these findings paint a clear picture. Borrowers with large businesses are substantially less likely to finish repaying their loans but the amount of money they take in default is low. One interpretation of these findings is that these richer customers no longer found it worthwhile to attend repayment meetings, either because they found the biweekly meetings too time consuming or because they no longer saw the value in maintaining the option for further microloans. While the borrowers who do not completely repay their loans are only a subset of those who do not continue to borrow, the pattern is highly suggestive. The primary loss to the MFI may come from losing these customers rather than the default itself, and presents the possibility that the MFI may want

\[14\text{This is the publicly available measure of profits.}\]
to restrict the business growth of their customers.

2.6.2 Comparative Statics on Likelihood of Restrictive Contracts

Among the most distinctive features of our model is the comparative static of contractual restrictions with respect to the level of competition. Specifically, Proposition 2.14 makes an unambiguous, testable prediction and in this section we present evidence in its support in the context of informal borrowing in Thai villages.

To do so we leverage variation induced by the 2002 Million Baht Program, which endowed each village with a village fund with one million baht to lend. Importantly, the amount of money endowed to each village fund is invariant to the village size - thus smaller villages received more credit per capita than larger villages. Kaboski and Townsend (2012) leverage this variation to estimate the return to credit offered by village funds. In contrast, we focus on informal money lenders and their borrowers, and study how this change in competition affected the contracts they offered. As we show below, there is a steep decline in the likelihood a villager borrows from a money lender as a function of the villager’s wealth, and thus we think of informal money lenders and their borrowers as an empirical analogue of the informal sector in our model.

In what follows, we present an empirical definition of contractual restrictions, and show that borrowers subjected to these restrictions indeed anticipate lower income the following year. Further, we show there is a negative correlation between the imposition of contractual restrictions and the interest rate, suggesting that money lenders may need to compensate borrowers whose investments they restrict. We next verify that the condition assumed by Proposition 2.14 is satisfied, and that as it predicts, money lenders in villages with larger increases in competition impose fewer contractual restrictions. Finally we invoke a variety of robustness and placebo tests to argue that this comparative static is not likely to be the result of other theories.

Data

The data used for this exercise were gathered as part of the ongoing Townsend Thai Project to track the financial lives of members of 64 Thai villages across 4 provinces: Chachoengsao, Lopburi, Sisaket, and Buriram. Specifically we utilize data from the household survey that tracked a representative sample of 15 households in each village on an annual basis. The dataset is extremely detailed; of particular relevance to this exercise it contains information about all loans received by study households (both formal and informal loans) including information on loan size, interest rates, collateral, consequence of default, and loan originator. The data also contain demographic information such as village size and composition, occupation, businesses operated, and a detailed breakdown of household income, and expectation about future income. For most of our regressions we utilize the unbalanced panel of loan level observations from survey rounds collected between
1997 (at the inception of the Townsend Thai Project) to 2007 (6 years after the Million Baht program was initiated).

**Validity of the Natural Experiment**

As described in Kaboski and Townsend (2012), two important elements make the Million Baht Program suitable for research. First, it was proposed during Prime Minister Thaksin Shinawatra’s election campaign following the dissolution of the Thai Parliament in November of 2000, and the program was then rapidly implemented between 2001 and 2002. So households were unlikely to anticipate the program in earlier years. Second the amount of credit endowed to each village was constant across villages of varying size, inducing variation in the per capita level of credit injected by the program. This motivates a continuous difference in differences strategy examining outcomes before and after the implementation of the program and across villages of varying sizes. That is, our principle regressions of interest will be of the form

$$outcome_{i,t} = \alpha + \beta_1 invsize_v + \beta_2 post_{i,t} + \beta_3 invsize_v \times post_{i,t} + \gamma X_{i,t} + \theta_t + \theta_h + \epsilon_{i,t}$$

where $invsize_v$ is the inverse size of the village, $post_{i,t}$ is an indicator variable that is 1 if year $t$ is after 2001, $X_{i,t}$ is a vector of loan and household characteristics for the borrower of loan $i$ in year $t$, $\theta_t$ is a year fixed effect, and $\theta_h$ is a household fixed effect. All standard errors are clustered at the village level.

The validity of this regression rests on the orthogonality assumption

$$post_{i,t}, invsize_v \perp \epsilon_{i,t}|X_{i,t}, \theta_t, \theta_h$$

Under the above assumption, $\beta_3$ captures the effect of treating small villages with a higher per capita credit shock than that of larger villages. To lend credibility to our identification assumption, we verify parallel trends in our pre-periods in the regressions to follow.

Further, in Table 1.2 we regress a number of village characteristics on inverse village size. We examine characteristics related to the village head, demographics, financial penetration, technology adoption, distance to a main road, and occupational distribution. The variables that have a significant relationship with inverse size are the number of agricultural cooperatives with a presence in the village, the education of the village head, and whether the village has common land that is shared among the villagers. In the regressions to follow we control for each of these characteristics as well as their interaction with $post_{i,t}$. Finally, as in Kaboski and Townsend (2012), our main analysis restricts attention to villages with between 50 and 250 households, but we show it is robust to including all villages.
Money Lenders are the Informal Sector

We focus on borrowers in these villages who receive loans from money lenders. The two defining features of informal lenders in the model are that richer villagers are less likely to borrow from them, and that they have some capacity to influence the project selection of their borrowers. In this section we argue that rich villagers stop borrowing from money lenders. Figure 1-2 plots the likelihood a household borrows from a money lender as a function of its wealth. As can be seen, the poorest households are about three times as likely to borrow from money lenders as are the richest households, creating the potential for money lenders to desire limiting the business growth of their clients.

Contractual Restrictions

Next we set forth two empirical definitions of contractual restrictiveness, one theoretically driven and one empirically driven. In line with our motivating examples our theoretical and primary measure of contractual restrictiveness is a binary indicator which takes a value of 1 if the borrower must forfeit productive capital for the duration of the loan or if the borrower must use a guarantor. In particular, in about 21% of cases in which land is used to collateralize a loan, the money lender is also given rights of use during the loan’s tenure. Because 60% of employed respondents report that farming is their primary means of income generation, and many more report it as a secondary means of income generation, we view this contractual feature as a restriction on the borrower’s ability to put a productive asset to use. Additionally, we code contracts which require a guarantor as restrictive, as guarantors may pressure borrowers to eschew profitable yet risky investments in favor of safer and less profitable ones to ensure repayment (see e.g. Banerjee et al. (1994) and Fischer (2013)).

All other forms of collateral, and the absence of a collateral requirement, are coded as unrestrictive. Other forms of collateral include the deed to the borrower’s land (but allow the borrower to retain the rights of use for the duration of the loan), jewelry, the deed to their house, and the right to proceed from future crop production. Importantly, none of these forms of collateral involve the transfer of productive capital from the borrower to the lender.

The defining feature of restrictive contracts in our model is that they restrict growth of the borrower’s business. Table 1.3 offers suggestive evidence that this may be the case in practice. Columns 1 and 2 present a regression of the borrower’s expected income in the following year on whether her money lender asks for a restrictive form of collateral. That is, we present regressions of the following form

\[ \log(y)_{i,t+1} = \alpha + \beta_{\text{restrictive}_{i,t}} + X_{i,t} + \theta_h + \theta_t + \epsilon_{i,t} \]
where \( \log(y)_{i,t+1} \) is the log of the borrower's year \( t \) expectation of her year \( t + 1 \) income, and all other variables are as defined above. In addition to household and year fixed effects, the controls \( X_{i,t} \) include the borrower's income and her loan size. As can be seen from the table, there is a negative and significant correlation between the borrower's expected income and the imposition of restrictive contractual features. In the next two columns we present a regression of the form

\[
\log(y)_{i,t+1} = \alpha + \sum_j \beta_j \text{collateral}_{j,i,t} + X_{i,t} + \theta_h + \theta_t + \epsilon_{i,t}
\]

where \( \text{collateral}_{j,i,t} \) is 1 if the borrower of loan \( i \) in year \( t \) was asked for a collateral of type \( \text{collateral}_j \). Several forms of collateral have large negative relationships with the borrower's expected income. Consistent with the theoretically motivated definition of contractual restrictiveness, these are that the borrower forfeits her land, that the lender asks for multiple guarantors, and a catch-all category labeled "other." Collateral in "other" are typically other forms of third party guarantees, such as using a third party's assets to guarantee the loan. This motivates a data driven definition of contractual restrictiveness that takes a value of 1 if the collateral takes any of the preceding three forms and 0 otherwise. In Table 1.15 we show that our main regressions are robust to using this empirically driven notion of restrictiveness rather than our primary one.

Finally we verify that, as predicted by our theory, restrictive forms of collateral are correlated with a reduction in the interest rate. Specifically we regress

\[
\text{MonthlyInterest}_{i,t} = \alpha + \beta_{\text{restrictive}_{i,t}} + X_{i,t} + \theta_h + \theta_t + \epsilon_{i,t}
\]

where \( \text{MonthlyInterest}_{i,t} \) is the monthly interest rate that the borrower is charged for the loan, and all other variables are as defined above. We present this regression in Table 1.4.

While one could imagine stories for the negative relationship between contractual restrictions and future expected income in which the arrow of causality points in the opposite direction (e.g. borrowers with low expectations about their future income are more tolerant of restrictive forms of collateral), it is a little tougher to do the same for the negative relationship between contractual restrictions and the interest rate. In particular, if contractual restrictions were a proxy for low quality borrowers, we would expect a positive relationship between the two variables. Thus, while acknowledging that the above relationships are merely correlational, we take them to be reassuringly consistent with the predictions of the model.

The Theory Predicts Increased Competition Corresponds to a Reduction in Restrictive Contracts

In general, the theory's prediction about how money lenders should react to increased competition is ambiguous. On the one hand, competition among lenders increases the borrower's bargaining
power, which is a force increasing the frequency of unrestrictive contracts. On the other hand, if the lender anticipates a sufficiently large increase competition for relatively rich borrowers, the lender may increase the frequency of restrictive contracts for poorer borrowers. Which of these two forces dominates is in general sensitive to parametric assumptions.

Taking the theory literally, however, Proposition 2.14 asserts that if in equilibrium the lender offers all borrowers (above a certain wealth level) a restrictive contract with strictly interior probability, the comparative static is unambiguous. In such a case we should expect that the first force always dominates, and that after the program was introduced, villages with a larger increase in competition (i.e. smaller villages) should have a correspondingly larger decrease in the frequency of restrictive contracts. Figure 1-3 plots the likelihood a borrower is asked for a restrictive form of collateral as a function of his income and verifies that the condition of Proposition 2.14 is satisfied. Thus the theory makes an unambiguous prediction.\textsuperscript{15}

Comparative Static on the Level of Competition

We now turn to our main regression:

\[
\text{restrictive}_{i,t} = \alpha + \beta_1\text{invsize}_v + \beta_2\text{post}_{i,t} + \beta_3\text{invsize}_v \times \text{post}_{i,t} + \gamma X_{i,t} + \theta_h + \theta_t + \epsilon_{i,t}
\]

where all variables are as defined above. In addition to household and year fixed effects, the controls include the three village characteristics correlated with inverse size and their interactions with \text{post}_{i,t}, the borrower’s income and loan size, and whether the household has a loan from the village fund. In our primary regressions we use data from 1997 to 2007, but also show in Table 1.14 that our results are robust to including only the three years before and after the program took effect. Note that our theoretical prediction is that \(\beta_3 < 0\), as higher inverse village size corresponds to a larger credit shock, which should correspond to a greater relaxation in contractual restrictiveness.

Our main results are presented in Table 1.5. As can be seen \(\beta_3\) has the desired sign and large magnitude. For instance, \(\beta_3\) is approximately \(-10.8\) in the OLS specification in column 1. The standard deviation of inverse village size is \(0.006\), so a one standard deviation increase in inverse village size corresponds to an approximately 6.5 percentage point decrease in the probability a restrictive form of collateral is demanded (off of a base of about 20 percentage points).

In Table 1.6, we verify parallel trends prior to 2002. Specifically, we regress

\[
\text{restrictive}_{i,t} = \alpha + \beta_1\text{invsize}_v + \beta_2\text{wave}_t + \beta_3\text{invsize}_v \times \text{wave}_t + \gamma X_{i,t} + \theta_h + \theta_t + \epsilon_{i,t}
\]

\textsuperscript{15}We do note that Proposition 2.14 requires the lender is following a mixed strategy between offering restrictive and unrestrictive contracts. While it is difficult to verify whether money lenders are indifferent between the two contracts in practice, it is reassuring that the true distribution of restrictive contracts is far from either corner case for all sufficiently wealthy borrowers.

151
and verify that we cannot reject $\beta_3 = 0$. We also regress

$$\text{restrictive}_{i,t} = \alpha + \beta_1 \text{inverse}_v + \sum_{t=1997}^{2001} \beta_t \text{wave}_t + \sum_{t=1997}^{2001} \beta_t \text{inverse}_v \times \text{wave}_t + \gamma X_{i,t} + \theta_h + \theta_t + \epsilon_{i,t}$$

and verify we cannot reject that $\tilde{\beta}_t$ are jointly 0.

Finally we regress

$$\text{restrictive}_{i,t} = \alpha + \beta_1 \text{inverse}_v + \sum_{t=1997}^{2007} \beta_t \text{wave}_t + \sum_{t=1997}^{2007} \beta_t \text{inverse}_v \times \text{wave}_t + \gamma X_{i,t} + \theta_h + \theta_t + \epsilon_{i,t}$$

and plot the coefficients $\beta_t$ in Figure 1-4. As can be seen, there does not seem to be a strong trend before or after 2002, but with the exception of 2005, all coefficients post 2002 are below 5 and all coefficients pre 2002 are above 5.

The Comparative Static is Not Driven by Borrower Selection

One threat to identification would be that the villagers who borrow from money lenders are selected differently in high and low competition environments. While our main regression would still capture the causal effect of competition on the incidence of restrictive contracts, it would be capturing an effect on the composition of borrowers rather than an effect on borrowers’ bargaining power. Because we use an unbalanced panel (we use loan level observations), our inclusion of household fixed effects does not eliminate the threat of selection at the household level.

Nevertheless we argue selection is unlikely to be driving our results. First we examine whether fewer households borrow from money lenders in high competition environments. Specifically we regress

$$\text{borrower}_{i,t} = \alpha + \beta_1 \text{inverse}_v + \beta_2 \text{post}_t + \beta_3 \text{inverse}_v \times \text{post}_t + \gamma X_{i,t} + \theta_h + \theta_t + \epsilon_{i,t}$$

where our unit of observation $i,t$ is a household $\times$ wave, our outcome $\text{borrower}_{i,t}$ is an indicator taking a value of 1 if household $i$ borrows from a money lender in wave $t$ and 0 otherwise, and everything else is defined as above. These regressions are presented in Table 1.7. The estimates fluctuate in sign, are always small in economic terms, and are all far from statistically significant. For example, recalling that standard deviation on inverse village size is .006, the estimates in column 2 imply that a village with one standard deviation smaller size saw a .2% larger decline in the frequency with which villagers borrow from money lenders after the Million Baht Program.

Next we examine whether households borrow less money from money lenders in villages with
higher competition. We regress

\[ \log(\text{amount})_{i,t} = \alpha + \beta_1 \text{invisize}_v + \beta_2 \text{post}_{i,t} + \beta_3 \text{invisize}_v \times \text{post}_{i,t} + \gamma X_{i,t} + \theta_h + \theta_t + \epsilon_{i,t} \]

where \( \log(\text{amount})_{i,t} \) is the log of the total amount borrowed from any money lender by household \( i \) in wave \( t \) and everything else is as defined above. The results are presented in Table 1.8. Again, the estimates fluctuate in sign, are always economically small, and far from statistically significant.

Finally Table 1.9 presents the results of a Heckman selection model of our primary specification. The selection equation assumes that the error terms are jointly normal and conditions on a number of village and borrower characteristics including \( \text{post}_{i,t}, \text{invisize}_v \) and \( \text{post}_{i,t} \times \text{invisize}_v \). Two observations are of note. First, the coefficient on interaction term \( \text{invisize}_v \times \text{post}_{i,t} \) in the selection equation is far from significant. Second, the estimates in Table 1.9 are strikingly similar to the corresponding estimates in Table 1.5. Taken together these offer further reassurance that selection is not driving our results.

**Ruling Out Alternative Theories**

We now present a number of tests to rule out alternative theories. First we aim to rule out a general trend away from restrictive forms of collateral. Specifically we rerun our main regression, but rather than focusing on the population of borrowers interacting with informal money lenders, we focus on the borrowers who take loans from their neighbors. These lenders should have no desire to keep their borrowers in poverty, and thus we would expect no relationship between the variation induced by the Million Baht Program and the likelihood that these lenders ask for restrictive forms of collateral. The regressions are presented in Table 1.10. The estimate on the interaction term fluctuates in sign and magnitude and is never statistically significant.

Next, we aim to separate our theory from any other one that makes no distinction between different forms of collateral. Again focusing on the population of borrowers who take loans from money lenders we regress

\[ \text{unrestrictive}_{i,t} = \alpha + \beta_1 \text{invisize}_v + \beta_2 \text{post}_{i,t} + \beta_3 \text{invisize}_v \times \text{post}_{i,t} + \gamma X_{i,t} + \theta_v + \theta_t + \epsilon_{i,t} \]

where \( \text{unrestrictive}_{i,t} \) is an indicator taking a value of 1 if a money lender asks for any form of collateral that was not coded as restrictive and 0 otherwise. We present this regression in Table 11 from which it is apparent that there is a positive relationship between the interaction term and the likelihood a money lender asks for unrestrictive collateral. This is what one would expect if money lenders substitute away from restrictive forms of collateral toward unrestrictive forms of collateral in high competition environments. Thus any theory consistent with these results must draw a distinction between restrictive and unrestrictive forms of collateral, and must also predict
the reduction in restrictive forms of collateral arising from increased competition from village funds.

Finally, we aim to rule out the possibility that the increased competition from village funds caused money lenders to screen their borrowers less carefully, and that this explains the shift away from restrictive forms of collateral in high competition environments. Though we do not observe screening efforts directly, we provide indirect evidence that this is not the case by examining how the interest rates charged by money lenders are influenced by competition. Specifically we regress

\[ \text{MonthlyInterest}_{i,t} = \alpha + \beta_1 \text{invsize}_v + \beta_2 \text{post}_{i,t} + \beta_3 \text{invsize}_v \times \text{post}_{i,t} + \gamma X_{i,t} + \theta_v + \theta_t + \epsilon_{i,t} \]

where all variables are as defined above. The results are in presented in Table 1.12. Though the estimates are noisy, and diminish as we include finer levels of fixed effects, it appears that if anything money lenders charge *higher* interest rates in higher competition environments.\(^{16}\) Thus, a theory based on screening effort should predict that money lenders increased their screening efforts in high competition environments to compensate for the more adversely selected population of borrowers, and would thus predict a corresponding increase in restrictive contracts, counter to what we observe.

2.7 Discussion

In this paper we explored a novel theory of informal lending, in which relatively rich borrowers cease lending from informal financiers and enter the formal sector. Each period the borrower and her informal lender bargain not only over the interest rate, but also over a contractual restriction that governs the borrower's ability to invest in business expansion. In contrast with earlier theories (e.g. Braverman and Srinivasan (1981) and Braverman and Stiglitz (1982)) we show that the borrower may indeed be caught in a debt trap when she cannot commit to share the benefits of entering the formal sector.

Our theory therefore reconciles the seemingly disparate facts that small scale entrepreneurs enjoy very high return to capital yet are unable to leverage microcredit and other forms of informal finance to realize those high returns. Moreover it offers an explanation for the robust findings that relatively wealthier business owners do enjoy high returns from microcredit and that on average the demand for microcredit is low.

The theory also offers nuanced predictions on comparative statics of the lending environment. Increasing the attractiveness of the formal sector improves the bargaining power of rich borrowers and hence increases their welfare and relaxes the poverty trap. However the same improvement may harm the welfare of poorer borrowers; anticipating that rich borrowers have improved bargaining

\(^{15}\)While it may at first seem unintuitive that an increase in competition is accompanied by an increase in the interest rate charged by money lenders, it is entirely consistent with our theory. We have documented that as competition rises money lenders offer more unrestricted contracts, which tend to carry higher interest rates than restrictive ones.\(^{16}\)
power, the lender tightens contractual restrictions on poor borrowers to prevent them from reaching higher levels of wealth and exploiting their improved position. Similarly, and counter to standard intuitions, increasing the borrower’s patience (and hence her value for business expansion) can make relatively poor borrowers worse off, and tighten the poverty trap. We finally showed that, while the model primarily studies a monopolist lender, similar effects can operate in contexts with imperfect competition, and derived intuitive comparative statics on the level of competition.

Then, studying money lenders in rural Thailand, we offered empirical evidence for the comparative static prediction on competition. Utilizing the Townsend Thai data and the plausibly exogenous shock to competition induced by the Million Baht Program, we found that Thai money lenders in environments with high competition reduced the frequency with which they imposed contractual restrictions more than money lenders in low competition environments. We argue that, because the same effects cannot be found for loans given by neighbors or for other forms of collateral, other theories are unlikely to explain this robust pattern.

In addition to the theories cited in the introduction, it is worth contrasting our theory with two other theories development economists seem to carry with them in the field. The first of these theories might sensibly be labeled as “blaming the borrower.” These are theories that allude to the argument that many borrowers are not natural entrepreneurs, and are primarily self employed for lack of being able to find steady wage work (see e.g. Schoar (2010)). While these theories enjoy some empirical support, they are at best a partial explanation of the problem as they are inconsistent with the large impacts of cash grants, cited in our introduction.

Second are the theories that assign blame to the lender for not having figured out the right lending contract. These theories are implicit in each of the experiments that evaluate local modifications to standard contracts (see e.g. Gine and Karlan (2014), Attanasio et al. (2015), and Carpena et al. (2013) on joint liability, Field et al. (2013) on repayment flexibility, and Feigenberg, Field, and Pande (2013) on meeting frequency). While many of these papers contribute substantially to our understanding of the ways in which microfinance operates, none have so far generated a lasting impact on the models that MFIs employ.

In contrast, ours is a theory that assumes that borrowers have the competence to grow their business, and that lenders are well aware of the constraints imposed on borrowers by the lending paradigm. Instead we focus on the rents that lenders enjoy from retaining customers, and the fact that customers who become sufficiently wealthy will find alternative forms of finance. Part of the value of this theory therefore may very well be its distance from the main lines of reasoning maintained by empirical researchers.
2.8 Appendix

2.8.1 Additional Extensions to the Model

The Borrower is Privately Informed About Her Outside Option

In this section we explore an extension in which the borrower maintains some private information about her outside option. In particular we augment the model such that the borrower's autarkic endowment is privately known. If she rejects the lender's contract she receives an endowment of $E_w + \nu_t$. Let $\nu_t \sim G$ be a random variable privately known to the borrower, and redrawn each period in an iid manner from some distribution $G$. Further, assume that if the borrower accepts the lender's contract, her endowment is still $E_w + T_w$. One way to understand this is that in the event that the borrower rejects the lender's contract, a relative will offer her a gift of size $\nu_t$, which she can allocate flexibly between her projects. We make the following additional assumption on the range of $G$ to simplify the discussion.

**Assumption 2.6.** Let $G$ have bounded support with minimum 0 and maximum $\bar{\nu}$ such that $\bar{\nu} < \frac{h_w}{d_w}$.

The above assumption guarantees that the borrower will accept any unrestrictive contract. However, if the lender offers the borrower a restrictive contract, he will now face a standard screening problem. Because he would like to extract the maximum acceptable amount of income, borrowers with unusually good outside options will reject his offer. This is encoded in the following proposition.

**Proposition 2.15.** The borrower never rejects an unrestrictive contract on the equilibrium path. However the borrower may reject restrictive contracts with positive probability.

**Proof.** See Appendix.

This intuitive result offers an explanation for the low takeup of microcredit contracts referenced in the introduction. Lenders who offer restrictive contracts to borrowers aim to extract all of the additional income generated by the loan, but in doing so sometimes are too demanding and therefore fail to attract the borrower. In contrast, lenders who offer unrestricted contracts necessarily leave the borrower with excess surplus, and therefore demand for these contracts is high.

The Borrower Flexibly Allocates a Fraction of Her Income

In this section we explore an extension to the model in which, even when subjected to contractual restrictions, the borrower maintains flexible control over a fraction of her income. In doing so we aim to show that our main result is qualitatively robust. Rather than finding that the borrower may remain in inefficiently small forever, we now find that having access to a lender may slow the borrower's growth relative to her autarkic rate.
Formally, the model is as in Section 1.2 but after accepting a contract \( (R, a) \), the borrower is free to invest a fraction \( s < 1 \) of her endowment flexibly, irrespective of the contractual restriction \( a \) the lender imposes. Thus we have weakened the lender’s ability to influence the borrower’s project choice. We maintain all other assumptions from Sections 1.2 and 1.3, and make the following observation.

**Proposition 2.16.** For sufficiently small \( s \), the lender may impose the contractual restriction \( C \) on the equilibrium path. In such an equilibrium the borrower reaches the formal sector in finite time, but will grow more slowly than she would have in autarky.

**Proof.** See Appendix.

### 2.8.2 Omitted Proofs

**Proof of Proposition 2.1**

In state \( n \) the borrower chooses her investment allocation \( i \) to maximize

\[
B_n^{\text{aut}} = \max_{i \leq E_n} q_n (E_n - i) dt + e^{-\rho dt} \left( \frac{1 - e^{-\frac{i}{\phi_n} dt}}{1 - e^{-\frac{1}{\phi_n} dt}} B_{n+1}^{\text{aut}} + e^{-\frac{i}{\phi_n} dt} B_n^{\text{aut}} \right)
\]

At the optimal level of \( i \), the borrower’s continuation utility can be rewritten as

\[
B_n^{\text{aut}} = \frac{q_n (E_n - i) dt + e^{-\rho dt} \left( \frac{1 - e^{-\frac{i}{\phi_n} dt}}{1 - e^{-\frac{1}{\phi_n} dt}} B_{n+1}^{\text{aut}} \right)}{1 - e^{-\left( \frac{1}{\phi_n} + \rho \right) dt}}
\]

\[
\rightarrow \frac{q_n (E_n - i)}{\frac{1}{\phi_n} + \rho} + \frac{\frac{1}{\phi_n}}{\frac{1}{\phi_n} + \rho} B_{n+1}^{\text{aut}}
\]

Because the problem is stationary, the borrower’s maximization problem can equivalently be written as choosing \( i \) to maximize her continuation utility above.

We now take the derivative of \( B_n^{\text{aut}} \) with respect to \( i \).

\[
\frac{dB_n^{\text{aut}}}{di} = \frac{-q_n \rho - q_n \kappa_n}{\left( \frac{1}{\phi_n} + \rho \right)^2} + \frac{\rho}{\phi_n} B_{n+1}^{\text{aut}}
\]

where recall \( \kappa_w \equiv \frac{E_w}{\phi_w} \). The denominator is positive. The numerator is positive iff

\[
-q_n \rho - q_n \kappa_n + \frac{\rho}{\phi_n} B_{n+1}^{\text{aut}} > 0
\]

\[
\iff \quad \alpha_n B_{n+1}^{\text{aut}} > \frac{q_n E_n}{\rho}
\]
We conclude that if \( \frac{q_n E_n}{p} < \alpha_n B^\text{aut}_{n+1} \) then the borrower's value in state \( n \) \( B^\text{aut}_n \) is increasing in \( i \) and otherwise it is decreasing. The result for earlier states follows similarly by backward induction. This completes our proof.

Proof of Proposition 2.2

The existence of an equilibrium follows standard arguments (See Maskin and Tirole (2001)). In this section we prove that generically the equilibrium is unique. We do so by backward induction on the state. In each state we first argue that if there exists an equilibrium in which the borrower invests her autarkic endowment in the working capital project conditional on rejecting the offered contract, then she does so in all equilibria and the lender gives her an unrestrictive contract with lower than necessary repayment rate.

We then argue that in any equilibrium where the borrower invests her autarkic endowment in the fixed capital project conditional on rejecting the offered contract (which, by the above statement, can only exist in the absence of an equilibrium in which the borrower invests her autarkic flow payoff in the working capital project), the lender's behavior is uniquely determined.

We first consider equilibrium behavior in state \( n \).

**Lemma 2.1.** There is no equilibrium in which the borrower weakly prefers to invest her autarkic endowment in the working capital project in state \( n \).

**Proof.** Suppose toward contradiction that upon rejecting the contract, the borrower weakly prefers investing her autarkic endowment in the working capital project. Then the borrower would accept any restrictive contract (which necessarily grants her more consumption than her outside option), and in state \( n \) the lender would clearly choose to give the borrower a maximally extractive restrictive contract in every period. The borrower's equilibrium continuation value is \( B_n = \frac{h_n + q_n E_n}{p} \) which by Assumption 2.3 is less than \( \alpha_n \frac{B}{p} \). This contradicts that the borrower weakly prefers to consume her autarkic flow payoff.

We next establish that conditional on the borrower strictly preferring to invest her autarkic endowment in fixed capital in equilibrium, the lender's behavior is uniquely determined.

**Lemma 2.2.** The probability the lender offers a restrictive contract in state \( n \), \( p_n \) is generically uniquely determined across any equilibrium in which the borrower strictly prefers to invest her autarkic endowment in the fixed capital project in state \( n \).

**Proof.** Assume the borrower strictly prefers to invest her autarkic endowment in the fixed capital project. Then if the lender offers the borrower any unrestrictive contract she accepts it, and the lender never benefits in state \( n \) from offering an excessively generous unrestrictive contract. Further, the lender never offers the borrower an excessively generous restrictive contract. So in equilibrium,
the lender either offers the borrower the contract \((q_n T_n - h_n, I)\), or the contract \((R(p), C)\) for some \(R(p)\) that pushes the borrower to her outside option. Now conjecture that in equilibrium the lender offers the borrower a restrictive contract with probability \(p\).

Noting that the borrower receives the maximum of the utility from investing her outside option in fixed capital and from consuming \(q_n E_n + h_n\) upon receiving a restrictive contract, we have

\[
B_n(p) = p \left( \max \left\{ e^{-\rho dt} \left( \left(1 - e^{-\kappa_n dt}\right) \frac{u}{\rho} + e^{-\kappa_n dt} B_n(p) \right), (q_n E_n + h_n) dt + e^{-\rho dt} B_n(p) \right\} \right)
+ (1 - p) \left( e^{-\rho dt} \left( \left(1 - e^{-\gamma dt}\right) \frac{u}{\rho} + e^{-\gamma dt} B_n(p) \right) \right)
\]

\[
B_n(p) = \max \left\{ p \left(1 - e^{-\kappa_n dt}\right) + (1 - p) \left(1 - e^{-\gamma dt}\right) e^{-\rho dt} \frac{u}{\rho} p (q_n E_n + h_n) dt + (1 - p) e^{-\rho dt} \left(1 - e^{-\gamma dt}\right) \frac{u}{\rho} \right\}
\]

It is straightforward to verify that \(\frac{dB_n(p)}{dp} < 0\). This is intuitive as a restrictive contract pushes the borrower to her individual rationality constraint (if possible), whereas an investment contract does not.

The highest possible repayment rate \(R(p)\) that can be required for a restrictive contract is pinned down by the borrower’s individual rationality constraint

\[
(q_n (E_n + T_n) - R(p)) dt + e^{-\rho dt} B_n(p) \geq e^{-\rho dt} \left( \left(1 - e^{-\kappa_n dt}\right) \frac{u}{\rho} + e^{-\kappa_n dt} B_n(p) \right)
\]

\[
\implies q_n (E_n + T_n) - R(p) = \max \left\{ \frac{e^{-\rho dt}}{dt} \left(1 - e^{-\kappa_n dt}\right) \left( \frac{u}{\rho} - B_n(p) \right), q_n E_n + h_n \right\}
\]

The maximal acceptable repayment rate is increasing in \(B_n(p)\) - this is intuitive as the higher is the borrower’s continuation value in state \(n\), the less she values investment.\(^{17}\)

Now, consider the lender’s decision of whether to offer an unrestricted or restrictive contract. Fixing the borrower’s expectation that the lender offers a restrictive contract with probability \(p\), in any period in which the lender offers an unrestricted contract his utility is

\[
(q_n T_n - h_n - T_n) dt + e^{-\rho dt} \left(1 - e^{-\gamma dt}\right) L_n(p)
\]

If he offers a restrictive contract his utility is

\[
(R(p) - T_n) dt + e^{-\rho dt} L_n(p)
\]

\(^{17}\)Note that by Assumption 2.3, \(q_n (E_n + T_n) - R(1) > q_n E_n + h_n\) for sufficiently small \(dt\), but in general \(q_n (E_n + T_n) - R(p)\) may equal \(q_n E_n + h_n\) for some \(p < 1\).
So he offers a restrictive contract if and only if the following incentive compatibility constraint holds

\[
(R(p) - T_n) \, dt + e^{-\rho dt} L_n(p) \geq (q_n T_n - h_n - T_n) \, dt + e^{-\rho dt} \left(1 - e^{-\gamma_n dt}\right) L_n(p)
\]

\[
\iff (q_n T_n - h_n - R(p)) \, dt \leq e^{-(\rho+\gamma_n)dt} L_n(p)
\]

The left hand side is the additional consumption the lender must forgo to persuade the borrower to accept a restrictive contract, and the right hand side is the discounted expected loss the lender incurs from allowing the borrower to invest in fixed capital.

Note that the lender's continuation utility \(L_n(p)\) is weakly decreasing in \(p\). This is so because the set of restrictive contracts the borrower will accept is decreasing in \(p\), while the set of unrestrictive contracts is unchanged.

Thus the left hand side of the above inequality is weakly increasing in \(p\), and the right hand side is weakly decreasing in \(p\). Given the lender's incentive compatibility constraint, we argue that generically there can only be one equilibrium level of \(p\).

If at \(p = 0\) (pure unrestrictive), the lender's incentive compatibility constraint for unrestrictive contracts is strictly satisfied, i.e.

\[
(q_n T_n - h_n - R(0)) \, dt > e^{-(\rho+\gamma_n)dt} L_n(0)
\]

then it will be strictly satisfied for all higher levels of \(p\), contradicting that any \(p > 0\) can be supported in equilibrium.

If at \(p = 1\) (pure restrictive) the lender's incentive compatibility constraint for restrictive contracts is strictly satisfied, i.e.

\[
(q_n T_n - h_n - R(1)) \, dt < e^{-(\rho+\gamma_n)dt} L_n(1)
\]

then it will be strictly satisfied for all lower levels of \(p\), contradicting that any \(p < 1\) can be supported in equilibrium.

If neither of the above inequalities holds even weakly then by the intermediate value theorem there will be a \(\bar{p}\) at which

\[
(q_n T_n - h_n - R(\bar{p})) \, dt = e^{-(\rho+\gamma_n)dt} L_n(\bar{p})
\]

Note that when the borrower believes she will receive a restrictive contract with probability \(\bar{p}\), the amount of consumption she demands when given a restrictive contract, \(q_n (E_n + T_n) - R(\bar{p})\), is
strictly larger than than $q_n E_n + h_n$ (the minimum feasible consumption the lender can leave the borrower). To see this, note that by assumption

$$(q_n T_n - h_n - R(0)) \, dt < e^{-(\rho + \gamma_n) \, dt} L_n(0).$$

Now, supposing that $q_n (E_n + T_n) - R(p) = q_n E_n + h_n$, we’d have $L_n(\bar{p}) = L_n(0)$ (because the borrower is willing to accept all feasible contracts in both cases), which would imply that $(q_n T_n - h_n - R(\bar{p})) \, dt < e^{-(\rho + \gamma_n) \, dt} L_n(\bar{p})$ and would contradict that the lender is indifferent between restrictive and unrestrictive contracts. Therefore we know that $q_n (E_n + T_n) - R(\bar{p}) = q_n E_n + h_n$ and thus $\frac{dR(\bar{p})}{dp} < 0$. At $p > \bar{p}$ the lender will strictly prefer investment loans and at $p < \bar{p}$ the lender will strictly prefer consumption loans, contradicting that any $p \neq \bar{p}$ can be supported in equilibrium.\(^{18}\)

So far we have argued that in state $n$ the borrower and lender’s behavior are uniquely determined across all equilibria. We now proceed to complete the proof by backward induction. Suppose in all states $\bar{w} \geq w + 1$ it has been shown that equilibrium behavior is generically unique.

We split the proof for state $w$ into the following two lemmas.

**Lemma 2.3.** If there exists an equilibrium in which the borrower weakly prefers to invest her autarkic endowment in the working capital project in state $w$, she does so in all equilibria. Moreover, the lender offers her the same (unrestrictive) contract across all equilibria.

**Proof.** Assume that in equilibrium the borrower weakly prefers to invest her autarkic endowment in working capital in the event of rejecting the lender’s contract in state $w$.

Then the borrower would accept a maximally extractive restrictive contract. But if the lender were to offer one in equilibrium, this would contradict that the borrower would invest her autarkic endowment in working capital. To see this, note that Assumption 2.3 guarantees

$$\alpha_w B_{w+1} > \frac{q_w E_w + h_w}{p}.$$ 

Thus the borrower would strictly prefer to invest her entire autarkic endowment $E_w$ in fixed capital than to accept the maximally extractive contract $(q_w T_w - h_w, C)$ if she expects that the lender will offer her that contract in all future periods in state $w$.

So in equilibrium the lender must weakly prefer to offer the contract $(q_w T_w - h_w, I)$ to the contract $(q_w T_w - h_w, C)$. Suppose in equilibrium the lender offers the borrower an unrestrictive contract for which the borrower’s IR constraint binds. Then in every period the borrower is indifferent between consuming $q_w E_w$ and accepting the lender’s contract. The borrower’s continuation

\(^{18}\)Note that since both $R(p)$ and $L(p)$ can both be written in terms of exogenous parameters of the model, it will hold generically that neither 2.8.1 nor 2.8.2 holds with equality.

161
utility is \( B_w = \frac{q_w F_w}{\rho} \), which, by Assumption 2.3, contradicts her desire to invest her autarkic endowment in working capital.

The final possibility, which we expand on below, is that in equilibrium the lender offers the borrower an unrestricted contract \((R, I)\) that is more generous than the borrower’s IR constraint demands. But if the borrower’s IR constraint doesn’t bind when offered \((R, I)\), then it must be that this is the lender’s unconstrained optimal contract and will continue to be feasible in any conjectured equilibrium in which the borrower strictly prefers to invest her autarkic endowment in fixed capital. Noting that, by assumption she weakly prefers to invest her autarkic flow income in working capital when offered the contract \((R, I)\), this contradicts the existence of an alternative equilibrium.

Formally, if the borrower weakly prefers to invest her autarkic endowment in working capital, then the lender solves

\[
\max_R \left( R - T_w \right) dt + e^{-\rho dt} \left( 1 - e^{-\frac{E_w + T_w - \frac{R}{\phi_w}}{\phi_w} dt} \right) L_{w+1} - e^{-\frac{E_w + T_w - \frac{R}{\phi_w}}{\phi_w} dt} L_w
\]

subject to

\[
e^{-\rho dt} \left( 1 - e^{-\frac{E_w + T_w - \frac{R}{\phi_w}}{\phi_w} dt} \right) B_{w+1} + e^{-\frac{E_w + T_w - \frac{R}{\phi_w}}{\phi_w} dt} B_w \geq q_w E_w dt + e^{-\rho dt} B_w
\]

where

\[
T_w \leq R \leq q_w T_w - h_w
\]

Note that, because the problem is stationary, we could equivalently solve for the contract that maximizes the lender’s continuation value in state \(w\). That is we can solve

\[
\max_R \left( R - T_w \right) dt + e^{-\rho dt} \left( 1 - e^{-\frac{E_w + T_w - \frac{R}{\phi_w}}{\phi_w} dt} \right) L_{w+1}
\]

subject to

\[
e^{-\rho dt} \left( 1 - e^{-\frac{E_w + T_w - \frac{R}{\phi_w}}{\phi_w} dt} \right) B_{w+1} + e^{-\frac{E_w + T_w - \frac{R}{\phi_w}}{\phi_w} dt} B_w \geq q_w E_w dt + e^{-\rho dt} B_w
\]

where

\[
T_w \leq R \leq q_w T_w - h_w
\]

It is straightforward to show that the maximand is increasing if and only if

\[
L_{w+1} > \frac{\phi_w}{q_w e^{-\rho dt}} \left[ 1 - e^{-\frac{E_w + T_w - \frac{R}{\phi_w}}{\phi_w} dt} \right] e^{-\frac{E_w + T_w - \frac{R}{\phi_w}}{\phi_w} dt} - \left( \rho + \frac{E_w + T_w - \frac{R}{\phi_w}}{\phi_w} dt \right)
\]

162
and that
$$\left(1 - e^{\frac{E_w + T_w - \frac{R}{\phi_w}}{\phi_w}}\right) dt$$
is increasing and always less than 1 for $R \leq q_w T_w - h_w$.

Thus the maximand is either monotonically decreasing, monotonically increasing, or increasing and then decreasing. If the borrower’s IR constraint doesn’t bind, the lender achieves a local maximum. And in any of these cases, the local maximum is also a global maximum in the range of repayments $R \in [T_w, q_w T_w - h_w]$. This uniquely pins down the repayment rate in an equilibrium in which the borrower invests her autarkic endowment in working capital.

Further, if such an equilibrium exists, there could not also be an equilibrium in which the borrower strictly preferred to invest her autarkic endowment in fixed capital. In such an equilibrium the lender would be solving

$$\max_R \left( (R - T_w) dt + e^{-\rho dt} \left(1 - e^{\frac{E_w + T_w - \frac{R}{\phi_w}}{\phi_w}}\right) L_{w+1} \right)$$

s.t.

$$e^{-\rho dt} \left(1 - e^{\frac{E_w + T_w - \frac{R}{\phi_w}}{\phi_w}}\right) B_{w+1} + e^{\frac{E_w + T_w - \frac{R}{\phi_w}}{\phi_w}} dt B_w \geq e^{-\rho dt} \left(1 - e^{\frac{E_w}{\phi_w}} dx\right) B_{w+1} + e^{\frac{E_w}{\phi_w}} dt B_w$$

where the borrower’s IR constraint clearly never binds. So in this case as well the lender achieves the global maximum within the range of repayments $R \in [T_w, q_w T_w - h_w]$, contradicting that the borrower would invest her autarkic endowment in fixed capital in the event that she has rejected the lender’s contract.

Thus if there exists an equilibrium where the borrower weakly prefers to invest her autarkic endowment in working capital in state $w$, she does so in all equilibria. We complete the proof by noting that in any equilibrium in which the borrower strictly prefers invest her autarkic endowment in fixed capital (which by the Lemma 2.3 only occurs when there is no equilibrium in which the borrower weakly prefers to invest her autarkic endowment in working capital), the lender’s behavior is uniquely pinned down.

**Lemma 2.4.** The probability the lender offers a restrictive contract in state $w$, $p_w$, is generically uniquely determined across any equilibrium in which the borrower strictly prefers to invest her autarkic endowment in the fixed capital project in state $w$. 

163
Proof. The proof proceeds exactly as in Lemma 2.2 and is thus omitted.

This completes the proof that the equilibrium is generically unique.

Proof of Proposition 2.3

We aim to show that in equilibrium the probability the lender offers the borrower a restrictive contract in state $w$, $p_w$, is single peaked in $w$. We split the proof into two steps. First we show that conditional on the borrower investing her autarkic endowment in fixed capital in a given region of states $w_k, \ldots, w_{k'}$ with $k' > k$, the probability the lender offers a restrictive contact is single peaked in the state.

Second we show that if in equilibrium the lender offers the borrower an excessively generous unrestricted contract in state $w$, then he also does so for all states $w' < w$, and hence $p_{w'} = 0$ in all states $w' < w$. Because we showed that in equilibrium the borrower only invests her outside option in working capital in states where she gets an excessively generous unrestricted contract, this completes the argument.

Lemma 2.5. Suppose that in equilibrium the borrower invests her autarkic endowment in fixed capital in states $\bar{w} - 1, \bar{w}$, and $\bar{w} + 1$. Then if $p_{\bar{w}} < p_{\bar{w}+1} \implies p_{\bar{w}-1} \leq p_{\bar{w}}$.

Proof. Assume that in equilibrium the borrower invests her outside option in fixed capital in states $\bar{w} - 1, \bar{w}$ and $\bar{w} + 1$. We begin by defining a function that implicitly determines the equilibrium probability $p_{\bar{w}}$ that the lender offers the borrower a restrictive contract. To do so we first determine the borrower’s value in state $\bar{w}$ as a function of the probability $p$ she expects a restrictive contract. This allows us to determine the maximal repayment rate $R(p)$ she is willing to accept for a restrictive contract given the probability $p$ she expects the lender to offer a restrictive contract. Finally, $R(p)$ allows us to determine the lender’s payoff from offering restrictive contracts, and by comparing this to his payoff from offering unrestricted contracts we pin down the equilibrium probability $p_{\bar{w}}$.

In state $\bar{w}$, if in equilibrium the borrower receives a restrictive contract with probability $p_{\bar{w}}$, her utility is

\[ B_{\bar{w}} (p_{\bar{w}}) = e^{\rho dt} \left( p_{\bar{w}} \left( (1 - e^{-\kappa dt}) B_{\bar{w}+1} + e^{-\kappa dt} B_{\bar{w}} \right) + (1 - p_{\bar{w}}) \left( (1 - e^{-\gamma dt}) B_{\bar{w}+1} + e^{-\gamma dt} B_{\bar{w}} \right) \right) \]

\[ B_{\bar{w}} (p_{\bar{w}}) = \frac{p_{\bar{w}} e^{\rho dt} - e^{-(\kappa + \rho) dt} B_{\bar{w}+1} + (1 - p_{\bar{w}}) e^{-(\gamma + \rho) dt} B_{\bar{w}+1}}{1 - p_{\bar{w}} e^{-(\rho + \kappa) dt} - (1 - p_{\bar{w}}) e^{-(\rho + \gamma) dt}} \cdot \frac{p_{\bar{w}} \kappa B_{\bar{w}+1} + (1 - p_{\bar{w}}) \gamma B_{\bar{w}+1}}{\rho + \rho \kappa + (1 - p_{\bar{w}}) \gamma} = \frac{\delta (p_{\bar{w}})}{\rho + \delta (p_{\bar{w}})} B_{\bar{w}+1} \]

164
where \( \delta(p_w) \equiv p_w \kappa + (1 - p_w) \gamma \).

Further recall that in equilibrium, the maximum repayment \( R(p) \) that the borrower would accept is given by

\[
(q_w (E + T_w) - R(p)) dt + e^{-\rho dt} B_w = e^{-\rho dt} \left( (1 - e^{-\kappa dt}) B_{w+1} + e^{-\kappa dt} B_w \right)
\]

\[
R(p) dt = q_w (E + T_w) dt - e^{-\rho dt} \left( (1 - e^{-\kappa dt}) (B_{w+1} - B_w) \right)
\]

Now, given the borrower’s equilibrium expectation \( p_w \), we can calculate the lender’s payoff from offering a maximally extractive, acceptable restrictive or unrestrictive contract. Because the problem is stationary, we can determine which contract the lender prefers by comparing the lender’s lifetime utility if he were to offer only restrictive contracts or only unrestrictive contracts in state \( \bar{w} \). If the lender were to offer only restrictive contracts in state \( \bar{w} \) his utility would be

\[
L^R_w(p_w) = (R(p_w) - T_{\bar{w}}) dt + e^{-\rho dt} L^R_w(p_w)
\]

\[
= (q_w (E + T_{\bar{w}}) - T_{\bar{w}}) dt + e^{-\rho dt} \left( (1 - e^{-\kappa dt}) (B_{\bar{w}+1} - B_{\bar{w}}) \right) + e^{-\rho dt} L^R_w(p_w)
\]

\[
= (q_w (E + T_{\bar{w}}) - T_{\bar{w}}) dt - e^{-\rho dt} \left( (1 - e^{-\kappa dt}) \left( \frac{\rho}{\rho + \delta(p_w)} B_{\bar{w}+1} \right) \right)
\]

\[
\rightarrow \frac{(q_w (E + T_{\bar{w}}) - T_{\bar{w}})}{\rho} - \frac{\kappa}{\rho + \delta(p_w)} B_{\bar{w}+1}
\]

On the other hand, if the lender were to offer only unrestrictive contracts in state \( \bar{w} \) his utility would be

\[
L^U_w(p_w) = (q_w T_{\bar{w}} - h - T_{\bar{w}}) dt + e^{-\rho dt} \left( (1 - e^{-\gamma dt}) L_{\bar{w}+1} + e^{-\gamma dt} L^U_w(p_w) \right)
\]

\[
= \frac{(q_w T_{\bar{w}} - h - T_{\bar{w}}) dt + \left( (e^{-\rho dt} - e^{-(\rho + \gamma) dt}) L_{\bar{w}+1} \right)}{1 - e^{-(\rho + \gamma) dt}}
\]

\[
\rightarrow \frac{(q_w T_{\bar{w}} - h - T_{\bar{w}}) + \gamma L_{\bar{w}+1}}{\rho + \gamma}
\]

\[
= \frac{(1 - \beta) (q_w T_{\bar{w}} - h - T_{\bar{w}})}{\rho} + \beta L_{\bar{w}+1}
\]

Next, consider the function \( g_w(p) \equiv L^U_w(p) - L^R_w(p) \). Note that \( L^U_w(p) \) is independent of \( p \) and \( L^R_w(p) \) is decreasing in \( p \), so \( g_w(p) \) is increasing. If \( g_w(1) < 0 \), then the unique equilibrium is for the lender to offer a restrictive contract with probability 1. If \( g_w(0) > 0 \), then the unique equilibrium is for the lender to offer an unrestrictive contract with probability 1. Else, as shown in Proposition 2.2, generically there is a unique \( p_w \in [0, 1] \) such that \( g_w(p_w) = 0 \), and the unique equilibrium is for the lender to offer a restrictive contract with probability \( p_w \).
We now verify that $p_w$ is single peaked in $w$ by considering the following five exhaustive cases:

1. $0 < p_{\theta} < p_{\theta+1} < 1$
2. $0 < p_{\theta} < p_{\theta+1} = 1$

In these cases we aim to verify that $p_{\theta-1} < p_{\theta}$

3. $0 = p_{\theta} < p_{\theta+1} < 1$
4. $0 = p_{\theta} < p_{\theta+1} = 1$

In this case we aim to verify that $p_{\theta-1} = p_{\theta} = 0$ and $g_{\theta-1}(0) > g_{\theta}(0)$

5. $0 = p_{\theta} = p_{\theta+1}$ and $g_{\theta}(0) > g_{\theta+1}(0)$

In this case we aim to verify that $p_{\theta-1} = p_{\theta} = 0$ and $g_{\theta-1}(0) > g_{\theta}(0)$.

We will provide the proof for the case where $0 < p_{\theta} < p_{\theta+1} < 1$ and omit the others as they are all similar.

Because $p_{\theta+1}$ is interior, we have

$$L_{\theta+1} = L^R_{\theta+1}(p_{\theta+1}) = \frac{q_{\theta+1}(E + T_{\theta+1}) - T_{\theta+1}}{\rho} - \frac{\kappa}{\rho + \delta(p_{\theta+1})} B_{\theta+2}$$

and

$$B_{\theta+1} = \frac{\delta(p_{\theta+1})}{p + \delta(p_{\theta+1})} B_{\theta+2}$$

Thus in state $w$ we have

$$g_{\theta}(p_{\theta}) = L^U_{\theta} - L^R_{\theta}(p_{\theta})$$

$$= (1 - \beta) \left( \frac{q_{\theta} (E + T_{\theta}) - h - T_{\theta}}{\rho} \right) + \beta \left( \frac{q_{\theta+1}(E + T_{\theta+1}) - T_{\theta+1}}{\rho} - \frac{\kappa}{\rho + \delta(p_{\theta+1})} B_{\theta+2} \right)$$

$$= \beta \left( \frac{q_{\theta+1}(E + T_{\theta+1}) - T_{\theta+1}}{\rho} - \frac{\kappa}{\rho + \delta(p_{\theta+1})} B_{\theta+2} \right)$$

$$= 0$$

Similarly we have

$$g_{\theta-1}(p) = L^U_{\theta-1} - L^R_{\theta-1}(p)$$

$$= \beta \left( \frac{q_{\theta} (E + T_{\theta}) - h - T_{\theta}}{\rho} \right) - \left( \frac{q_{\theta-1}(E + T_{\theta-1}) - T_{\theta-1}}{\rho} - (1 - \beta) \frac{h + q_{\theta-1} E}{\rho} - \frac{\kappa}{\rho + \delta(p_{\theta-1})} B_{\theta+2} \right)$$

$$= 0$$

166
Comparing the expression for $g_{\hat{w}-1}(p)$ to the that of $g_\hat{w}(p)$ we can see that the sum of first two terms is strictly larger in the expression for $g_{\hat{w}-1}(p)$ (by Assumption 2.4). That means that in order for $p$ to set $g_{\hat{w}-1}(p) = 0$ (if possible), we need that the third term in $g_{\hat{w}-1}(p)$ is strictly smaller than the third term in $g_\hat{w}(p)$. That is

$$-\frac{\kappa}{\rho + \delta(p_\hat{w})} \frac{\beta - \delta(p_\hat{w})}{\rho + \delta(p_\hat{w})} B_{\hat{w}+2} > -\frac{\kappa}{\rho + \delta(p_\hat{w}+1)} B_{\hat{w}+2}$$

Recall that $p_{\hat{w}+1} > p_\hat{w}$ by assumption. So $\frac{\delta(p_{\hat{w}+1})}{\rho + \delta(p_\hat{w})} < 1$. Thus

$$\left(\beta - \frac{\delta(p_\hat{w})}{\rho + \delta(p)}\right) < \left(\beta - \frac{\delta(p_{\hat{w}+1})}{\rho + \delta(p_\hat{w})}\right)$$

On the other hand, if there is no $p \geq 0$ such that $g_{\hat{w}-1}(p) = 0$, then $g_{\hat{w}-1}(0) > 0$, and the unique equilibrium includes $p_{\hat{w}-1} = 0$. This completes the argument for this case. As the remaining cases are similar they are omitted.

To complete the proof we now only need to address the possibility that $p_\hat{w} = 0$ and that the lender offers the borrower an unrestrictive contract with smaller than necessary repayment in state $\hat{w}$ in equilibrium. We do so with the following lemma.

**Lemma 2.6.** Suppose $p_\hat{w} = 0$ and the lender offers the borrower an unrestrictive contract with smaller than necessary repayment in state $\hat{w}$ in equilibrium. Then for all $w < \hat{w}$, $p_w = 0$ and the lender gives the borrower an unrestrictive contract with a smaller than necessary repayment.

**Proof.** Given that the lender offers an unrestrictive contract in state $\hat{w}$, he optimally sets the repayment rate to maximize

$$L^U_{\hat{w}} = \max_{T_\hat{w} \leq R \leq q_\hat{w}T_\hat{w} - h} \left( R - T_\hat{w} \right) + \frac{E + T_\hat{w} - \frac{R}{\phi}}{\rho + \frac{E + T_\hat{w} - \frac{R}{\phi}}{q_\hat{w}}} L_{\hat{w}+1}$$

It is easily verified that the above objective function is monotonically decreasing (increasing) in $R$ if and only if $L_{\hat{w}+1} \geq \left(\frac{q_\hat{w}(E + T_\hat{w}) - T_\hat{w}}{\rho}\right) + \hat{w}$. That in state $\hat{w}$ the lender optimally offers a more generous than necessary repayment rate implies that $L_{\hat{w}+1} \geq \frac{q_\hat{w}(E + T_\hat{w}) - T_\hat{w}}{\rho} + \hat{w}$.
Now consider state \( \bar{w} - 1 \), and suppose toward contradiction that the lender does not offer an excessively generous unrestrictive contract in \( \bar{w} - 1 \). By the discussion above, this implies that

\[
L_{\bar{w}} \leq \frac{q_{\bar{w}-1}(E + T_{\bar{w}-1}) - T_{\bar{w}-1}}{\rho} + \phi \tag{2.8.3}
\]

Define \( w_R \) to be the lowest state larger than \( \bar{w} \) in which a restrictive contract is offered with positive probability (if no such state exists, the proposition holds trivially). We will show equation 2.8.3 implies that for all states \( w \in \{ \bar{w} - 1, \ldots w_R - 1 \} \) the lender offers the least generous unrestrictive contract, contradicting the premise of this lemma.

We know that

\[
L_{\bar{w}} \geq (1 - \beta) \left( \sum_{w' = \bar{w}}^{w_R-1} \frac{q_{w'} T_{w'} - h - T_{w'} \beta^{w' - \bar{w}}}{\rho} \right) + \beta^{w_R - \bar{w}} L_{w_R},
\]

as the lender would derive this utility if he allowed the borrower to invest in fixed capital at the slowest possible rate in each state \( w' > \bar{w} \) until \( w_R \) (by assumption he optimally allows the borrower to invest at a weakly higher rate).

Combining the above two inequalities we have that

\[
(1 - \beta) \left( \sum_{w' = \bar{w}}^{w_R-1} \frac{q_{w'} T_{w'} - h - T_{w'} \beta^{w' - \bar{w}}}{\rho} \right) + \beta^{w_R - \bar{w}} L_{w_R} \leq \frac{q_{\bar{w}-1}(E + T_{\bar{w}-1}) - T_{\bar{w}-1}}{\rho} + \phi \tag{2.8.4}
\]

We now show that the above inequality together with concavity of \( \frac{q_w(E + T_w) - T_w}{\rho} \) in \( w \) (Assumption 2.4) implies

\[
(1 - \beta) \left( \sum_{w' = \bar{w}+1}^{w_R-1} \frac{q_{w'} T_{w'} - h - T_{w'} \beta^{w' - (\bar{w}+1)}}{\rho} \right) + \beta^{w_R - (\bar{w}+1)} L_{w_R} \leq \frac{q_{\bar{w}}(E + T_{\bar{w}}) - T_{\bar{w}}}{\rho} + \phi \tag{2.8.5}
\]

The right hand side of equation 2.8.5 is \( \frac{q_{\bar{w}}(E + T_{\bar{w}}) - T_{\bar{w}}}{\rho} - \frac{q_{\bar{w}-1}(E + T_{\bar{w}-1}) - T_{\bar{w}-1}}{\rho} \) larger than that of 2.8.4. To evaluate the difference in left hand sides note
The first equality follows from rearrangement of terms, the second follows from substituting for $L_{wR}$, the third follows from further rearrangement, the fourth inequality follows from deletion of a negative term, and the fifth inequality follows from the concavity of $q_w(E + Tw) - Tw$ in $w$. Thus the difference in the left hand side of equations 2.8.5 and 2.8.4 is smaller than that of the right hand sides. So we conclude that equation 2.8.5 holds.

Proceeding by forward induction we have that

$$
(1 - \beta) \left( \sum_{w'=w+1}^{w_R-1} \frac{q_w T_{w'} - h - T_{w'} \beta^{w'-(w+2)}}{\rho} \right) + \beta^{w_R-(w+1)} L_{wR} \leq \frac{q_w (E + T_w) - T_w}{\rho} + \phi
$$

for all $\tilde{w}$ such that $\tilde{w} \leq w \leq w_R - 1$. Note that at $w = w_R - 1$ the above equation implies that $L_{wR} \leq \frac{q_{wR-1}(E + T_{wR-1}) - T_{wR-1}}{\rho} + \phi$ and hence at $w_R - 1$ the lender invests at the lowest possible rate. Therefore $L_{wR-1} = (1 - \beta) \frac{q_{wR-1} T_{wR-1} - h - T_{wR-1}}{\rho} + \beta L_{wR}$. Proceeding backward the argument extends for all $w$ such that $\tilde{w} \leq w \leq w_R - 1$, so that $L_w = (1 - \beta) \left( \sum_{w'=w}^{w_R-1} \frac{q_w T_{w'} - h - T_{w'} \beta^{w'-w}}{\rho} \right) + \beta^{wR-w} L_{wR}$ for $\tilde{w} \leq w \leq w_R - 1$ and hence the lender allows the borrower to invests at the lowest possible rate in all such states. This completes the contradiction.

Together the above two lemmas complete the proof.
Proof of Proposition 2.4

Recall from the proof of Proposition 2.2 that in equilibrium the lender offers a restrictive contract in state $w$ with probability 1 if and only if $L^R_w (1) \geq L^U_w$, where $L^R_w (p) = \frac{q_w (E+T_w) - T_w}{\rho} - \frac{\kappa}{\rho + \delta (p)} B_{w+1}$ and $L^U_w = (1 - \beta) \frac{q_w T_w - h - T_w}{\rho} + \beta L_{w+1}$. Now

$$L^R_w (1) \geq L^U_w$$

$$\iff \frac{q_w (E + T_w) - T_w}{\rho} - \alpha B_{w+1} > (1 - \beta) \frac{q_w T_w - h - T_w}{\rho} + \beta L_{w+1}$$

$$\iff -\alpha B_{w+1} > -(1 - \beta) \left( \frac{q_w E + h}{\rho} \right) + \beta \left( L_{w+1} - \frac{q_w (E + T_w) - T_w}{\rho} \right)$$

$$\iff (\beta - \alpha) B_{w+1} > \beta \left( B_{w+1} + L_{w+1} - \frac{q_w (E + T_w) - T_w}{\rho} \phi \right)$$

which completes the proof.

Proof of Proposition 2.5

Suppose $h = 0$ and consider the lender's behavior in state $n$. Fixing a probability $p_n$ that the borrower anticipates a restrictive contract in equilibrium, (and recalling that we can consider the lender's continuation utility in state $n$ from a fixed action due to the stationarity of the problem), the lender's utility from offering a restrictive contract is

$$L^R_n (p_n) = \frac{q_n (E + T_n) - T_n}{\rho} - \frac{\kappa}{\rho + \delta (p_n)} u$$

$$= \frac{q_n (E + T_n) - T_n}{\rho} - \frac{u}{\alpha}$$

where the equality follows from the fact that $h = 0$.

On the other hand, his utility from offering an unrestricted contract is

$$L^U_n (p_n) = \frac{q_n T_n - h - T_n}{\rho} (1 - \beta)$$

$$= \frac{q_n T_n - h - T_n}{\rho} (1 - \alpha)$$

170
The lender prefers offering an unrestrictive contract over a restrictive one if and only if

\[
q_n T_n - h - T_n \frac{(1 - \alpha)}{\rho} \geq \frac{q_n (E + T_n) - T_n}{\rho} - \frac{\alpha u}{\rho}
\]

\[
(1 - \alpha) \frac{q_n T_n - h - T_n}{\rho} + \frac{\alpha u}{\rho} \geq \frac{q_n (E + T_n) - T_n}{\rho}
\]  

(2.8.6)

The left hand side of inequality 2.8.6 is the sum of the borrower and lender’s welfares if the borrower invests in fixed capital at the slowest possible rate in the relationship, and the right hand side is the sum of their welfares from if the borrower invests in working capital. Thus if it is socially efficient to invest in fixed capital the lender strictly prefers unrestrictive contracts, irrespective of the borrower’s expectation \( p_n \), and thus in equilibrium in state \( n \) the lender chooses unrestrictive contracts with probability 1.

Moving backwards, the proof proceeds similarly.

**Proof of Propositions 2.6 and 2.7**

**Lemma 2.7.** For any state \( w > \tilde{w} \), \( \frac{dp_w}{du} \leq 0 \) with strict inequality if \( p_w > 0 \).

Proof. By definition, in states \( w > \tilde{w} \), \( p_w < 1 \). Thus in equilibrium the lender at least weakly prefers offering the borrower an unrestrictive contract. We can thus write the lender’s continuation utility in each such state as the utility he derives from offering an unrestrictive contract at every period (fixing the borrower’s expectation at \( p_w \)). That is

\[
L_w = L^U_w \equiv (1 - \beta) \frac{q_w T_w - h - T_w}{\rho} + \beta L_{w+1}
\]

On the other hand, if the lender were to offer a minimally generous restrictive contract at every period (again, fixing the borrower’s expectation at \( p_w \)) he would receive a continuation utility of

\[
L_w = L^R_w (p_w) \equiv \frac{q_w (E + T_w) - T_w}{\rho} - \frac{\kappa}{\rho + \delta (p_w)} B_{w+1}
\]

In state \( n \) \( L^U_n = (1 - \beta) \frac{q_n T_n - h - T_n}{\rho} \) which is not a function of \( u \). On the other hand \( L^R_n (p_n) = \frac{q_n (E + T_n) - T_n}{\rho} - \frac{\kappa}{\rho + \delta (p_n)} \) is decreasing in \( u \). Hence if in state \( n \) \( L^U_n > L^R_n (0) \), then \( p_n = 0 \) and \( \frac{dp_n}{du} = 0 \). The lender’s continuation utility is \( L_n = (1 - \beta) \frac{q_n T_n - h - T_n}{\rho} \) and \( \frac{dp_n}{du} = 0 \). The borrower’s utility is \( B_n = \beta \frac{u}{\rho} \) so \( \frac{dp_n}{du} > 0 \).

Note that in full generality he may offer the borrower an unrestrictive contract with positive transfer in state \( w \). If so his continuation utility is \( L^U_w = \left( 1 - \frac{q_w (E + T_w) - R}{\alpha} \right) \frac{R - T_w}{\rho} + \left( \frac{q_w (E + T_w) - R}{\alpha + \rho} \right) L_{w+1} \), but otherwise the argument goes through unchanged.

171
If \( L_n^I = L_n^C(p_n) \) for some \( p_n \in [0, 1] \) then \( p_n \) is the solution to
\[
g(p_n) \equiv (1 - \beta) \frac{q_n T_n - h - T_n}{\rho} - \left( \frac{q_n (E + T_n) - T_n}{\rho} - \frac{\kappa}{\rho + \delta(p_n)} u \right) = 0
\]
Since \( \delta \) is a decreasing function, it is clear that \( \frac{dp_n}{du} < 0 \). But we still have \( L_n = (1 - \beta) \frac{q_n T_n - h - T_n}{\rho} \) so that \( \frac{dL_n}{du} = 0 \). \( B_n = \frac{\delta(p_n) u}{\rho + \delta(p_n) \rho}, \) so
\[
\frac{dB_n}{du} = \frac{(\frac{d\delta(p_n)}{du} \frac{dp_n}{du}) \rho u}{(\rho + \delta(p_n))^2} + \frac{\delta(p_n)}{\rho + \delta(p_n) > 0}
\]
Proceeding backward to any state \( w > \bar{w} \), suppose \( \frac{dB_{w+1}}{du} > 0, \frac{dL_{w+1}}{du} = 0. \) Then the proof proceeds exactly as above. This completes the proof of the lemma.

We next consider the comparative static in states \( w = \{w, ..., \bar{w}\} \).

**Lemma 2.8.** For \( w \in \{w, ..., \bar{w}\} \), generically \( \frac{dp_w}{du} = 0, \frac{dB_w}{du} > 0 \) and \( \frac{dL_w}{du} < 0 \).

**Proof.** By definition \( p_w = 1 \) for \( w \in \{w, ..., \bar{w}\} \). Generically this preference will be strict and thus \( \frac{dp_w}{du} = 0. \)

Recall that in Lemma 2.7 we established \( \frac{dB_{w+1}}{du} > 0. \) We also know that \( L_{\bar{w}} = L_{\bar{w}}^R(1) \equiv \frac{q_{\bar{w}} (E + T_{\bar{w}}) - T_{\bar{w}}}{\rho} - \alpha B_{\bar{w}+1}. \) Hence generically \( \frac{dL_{\bar{w}}}{du} < 0. \) Further, \( B_{\bar{w}} = \alpha B_{\bar{w}+1} \) so \( \frac{dB_{\bar{w}}}{du} = \alpha \frac{dB_{\bar{w}+1}}{du} > 0. \)

For the remainder of the states \( w \in \{w, ..., \bar{w}\} \), the result follows from straightforward induction.

We now consider the comparative statics for \( w < \bar{w} \) in the following three lemmas.

**Lemma 2.9.** Suppose \( p_{\bar{w}-1} = 0. \) Then generically \( \frac{dp_{\bar{w}}}{du} = 0, \frac{dL_{\bar{w}}}{du} < 0, \) and \( \frac{dB_{\bar{w}}}{du} > 0 \) for all \( w < \bar{w}. \)

**Proof.** If \( p_{\bar{w}-1} = 0 \) and \( L_{\bar{w}-1}^U > L_{\bar{w}-1}^R(0) \), then the lender’s preference for unrestricted contracts is strict so \( \frac{dp_{\bar{w}-1}}{du} = 0. \) Further, Lemma 2.8 established that \( \frac{dL_{\bar{w}}}{du} < 0 \) and \( \frac{dB_{\bar{w}}}{du} > 0. \) Therefore, because \( L_{\bar{w}-1} = (1 - \beta) \frac{q_{\bar{w}-1} T_{\bar{w}-1} - h}{\rho} + \beta L_{w} \), we know \( \frac{dL_{\bar{w}-1}}{du} < 0. \) And \( B_{\bar{w}-1} = \beta B_{\bar{w}} \) so \( \frac{dB_{\bar{w}-1}}{du} > 0. \) Moving backwards proceeds by straightforward induction.

The remainder of the proof deals with the case for which \( p_{\bar{w}-1} > 0. \) We split the analysis into two cases based on the players’ level of patience.

**Lemma 2.10.** Suppose \( p_{\bar{w}-1} > 0 \) and \( \rho > \frac{\kappa T}{\kappa + \gamma}. \) Then \( \frac{dp_{\bar{w}}}{du} > 0, \frac{dL_{\bar{w}}}{du} < 0, \) and \( \frac{dB_{\bar{w}}}{du} < 0 \) for all \( w < \bar{w}. \)
Proof. Consider first state \( w - 1 \). We know \( p_{w-1} \) is the solution to \( g(p_{w-1}) = 0 \). So

\[
\beta \frac{(qw(E + Tw) - Tw)}{\rho} - \frac{(qw-1(E + T_{w-1}) - T_{w-1})}{\rho} - (1 - \beta) \frac{h + qw_{-1}E}{\rho} = \left( \beta - \frac{\delta(p_{w})}{\rho + \delta(p_{w})} \right) \frac{\kappa}{\rho + \delta(p_{w})} B_{w+1} = 0
\]

\[
\Longleftrightarrow
\beta \frac{(qw(E + Tw) - Tw)}{\rho} - \frac{(qw-1(E + T_{w-1}) - T_{w-1})}{\rho} - (1 - \beta) \frac{h + qw_{-1}E}{\rho} = \left( \beta - \frac{\delta(p_{w})}{\rho + \delta(p_{w})} \right) \frac{\kappa}{\rho + \delta(p_{w})} B_{w+1}
\]

Note that the left hand side of the above equation is constant in \( u \).

Thus

\[
\frac{d}{du} \left( \beta \frac{\kappa}{\rho + \delta(p_{w})} - \frac{\kappa}{\rho + \delta(p_{w-1})} \frac{\delta(p_{w})}{\rho + \delta(p_{w})} \right) B_{w+1} = 0 \quad (2.8.7)
\]

\[
\Longleftrightarrow
\frac{\kappa d\delta(p_{w})}{dp_{w}} \frac{dp_{w}}{du} + \frac{\kappa d\delta(p_{w-1})}{dp_{w-1}} \frac{dp_{w-1}}{du} - \delta(p_{w}) \frac{\kappa d\delta(p_{w})}{dp_{w}} \frac{dp_{w}}{du} - \delta(p_{w-1}) \frac{\kappa d\delta(p_{w-1})}{dp_{w-1}} \frac{dp_{w-1}}{du} = 0
\]

\[
\Longleftrightarrow
\frac{\kappa d\delta(p_{w})}{dp_{w}} \frac{dp_{w}}{du} - \delta(p_{w}) \frac{\kappa d\delta(p_{w})}{dp_{w}} \frac{dp_{w}}{du} = 0
\]

\[
\Longleftrightarrow
\frac{\kappa d\delta(p_{w-1})}{dp_{w-1}} \frac{dp_{w-1}}{du} < 0
\]

Where the second implication follows by removing positive terms from the right hand side and noting that \( p_{w} > p_{w-1} \) which implies that \( \frac{\delta(p_{w})}{\rho + \delta(p_{w-1})} < \frac{\delta(p_{w-1})}{\rho + \delta(p_{w-1})} \leq \beta \).

Next, note that

\[
L_{w-1} = L_{w-1}' = (1 - \beta) \frac{qw_{-1}T_{w-1} - h}{\rho} + \beta L_{w}
\]

so by Lemma 2.8 we know that \( \frac{dL_{w-1}}{du} < 0 \).

---

\( ^{20} \) The right hand side of the above equation can be simplified by noting that \( p_{w} = 1 \), but we leave it in this more general form to economize on notation in the backward induction step.

\( ^{21} \) Note that in full generality the lender may offer the borrower an investment loan with positive transfer in state.
To find the sign of $\frac{dB_{W-1}}{du}$ recall that

$$\frac{p+\delta(p_{W-1})}{\delta(p_{W-1})} B_{W-1} = B_w = \frac{\delta(p_w)}{\rho+\delta(p_w)} B_{W+1}. $$

Hence we know that

$$d \left( \frac{\beta \frac{\kappa}{\rho+\delta(p_w)} - \frac{\kappa}{\rho+\delta(p_{W-1})}}{\rho+\delta(p_{W-1})} \right) B_{W+1} = 0$$

$$d \left( \frac{\beta \frac{\kappa}{\delta(p_w)} - \frac{\kappa}{\delta(p_{W-1})}}{\delta(p_{W-1})} \right) B_{W-1} = (2.8.9)$$

$$d \left( \beta + \frac{\beta\rho - \kappa}{\delta(p_{W-1})} \right) B_{W-1} =$$

$$\frac{- (\beta \rho - \kappa) \frac{\delta(p_{W-1})}{dp_{W-1}} \frac{dp_{W-1}}{du}}{\delta(p_{W-1})^2} B_{W-1} + \frac{dB_{W-1}}{du} \left( \beta + \frac{\beta \rho - \kappa}{\delta(p_{W-1})} \right)$$

where the second equality follows from noting that $p_w = 1$. Reducing we have

$$\frac{dB_{W-1}}{du} = NEG (\beta \rho - \kappa)$$

where $NEG$ is a negative constant. The one subtle algebraic reduction in this final step is that

$$\left( \beta + \frac{\beta \rho - \kappa}{\delta(p_{W-1})} \right) \left[ \beta - \frac{\kappa}{\rho+\delta(p_{W-1})} \right] > 0.$$

Since we have assumed $\rho > \frac{\kappa\gamma}{\kappa + \gamma}$, which is equivalent to $\rho \beta > \kappa$ we have $\frac{dB_{W-1}}{du} < 0$.

Moving backward to state $W-2$, suppose $p_{W-2} > 0$ (or $p_{W-2} = 0$, but $L_{W-2}^T - L_{W-2}^C(0) = 0$). Then $p_{W-2}$ is the solution to $g_{W-2} (p_{W-2}) = 0$. That is

$$(1 - \beta) \frac{q_{W-2} T_{W-2} - h}{\rho} + \beta L_{W-1} - \left( \frac{q_{W-2}(E + T_{W-2})}{\rho} - \frac{\kappa}{\rho + \delta(p_{W-2})} B_{W-1} \right) = 0$$

Differentiating both sides with respect to $u$ we see

$$\beta \frac{dL_{W-1}}{du} + \left( -\frac{\delta(p_{W-2})}{\rho+\delta(p_{W-2})} \frac{dp_{W-2}}{du} \right) \left( \frac{-\delta(p_{W-2})}{\rho+\delta(p_{W-2})} \frac{dp_{W-2}}{du} \right) B_{W-1} + \frac{\kappa}{\rho + \delta(p_{W-2})} \frac{dB_{W-1}}{du} = 0$$

(2.8.11)

If so his continuation utility is $L_{W-1}^U = \left( 1 - \frac{\delta(p_{W-1})}{\rho + \delta(p_{W-1})} \frac{R - T_{W-1}}{\rho} \right) L_w$, and the interest rate becomes weakly higher but otherwise the argument to follow goes through unchanged.
We know that \( \frac{dL_{W-1}}{du} < 0 \) and \( \frac{dB_{W-1}}{du} < 0 \). Hence \( \frac{dpw_{-2}}{du} > 0 \). Further

\[
L_{W-2} = (1 - \beta) \frac{qw_{-2}T_{w-2} - h - T_{w-2}}{\rho} + \beta L_{W-1}
\]

so \( \frac{dL_{W-2}}{du} < 0 \). And \( B_{W-2} = \frac{\delta(pw_{-2})}{\rho + \delta(pw_{-2})} B_{W-1} \) so

\[
\frac{dB_{W-2}}{du} = \frac{\delta(pw_{-2}) \frac{dpw_{-2}}{du}}{\rho + \delta(pw_{-2})} B_{W-1} + \frac{\delta(pw_{-2})}{\rho + \delta(pw_{-2})} \frac{dB_{W-1}}{du} < 0 \tag{2.8.12}
\]

If instead we had \( pw_{-2} = 0 \) and \( L_{W-2} > L_{W-2}(0) \), then \( \frac{dpw_{-2}}{du} = 0 \). \( L_{W-2} = (1 - \beta) \frac{qw_{-2}T_{w-2} - h}{\rho} + \beta L_{W-1} \) so \( \frac{dL_{W-2}}{du} < 0 \). \( B_{W-2} = \beta B_{W-1} \) so \( \frac{dB_{W-2}}{du} < 0 \).

Because we used only that \( \frac{dB_{W-1}}{du} < 0 \) and \( \frac{dL_{W-1}}{du} < 0 \), moving backwards from state \( w - 2 \) to state 0 is straightforward induction. \( \square \)

We now complete the proof of Propositions 2.6 and 2.7 by considering a patient borrower.

**Lemma 2.11.** Suppose \( pw_{-1} > 0 \) and \( \rho < \frac{\kappa}{\kappa + \gamma} \). Then \( \frac{dpw}{du} > 0 \) and \( \frac{dL_{W}}{du} < 0 \) for all \( w < w \).

**Proof.** In state \( w - 1 \) everything follows as it did in Lemma 2.10 except that \( \frac{dB_{w}}{du} > 0 \), determined by equation (2.8.10) is positive. In state \( w - 2 \), the considerations are similar. \( \frac{dpw_{-2}}{du} > 0 \) is determined by equation (2.8.7) (reducing all indices by 1) and \( \frac{dL_{w-2}}{du} < 0 \) is determined by (2.8.8) (reducing all indices by 1). However the sign of \( \frac{dB_{w-2}}{du} \), determined by (2.8.9) is now ambiguous.

Moving back to an arbitrary state \( w < w \) such that \( \frac{dB_{w}}{du} > 0 \), the considerations will be exactly the same as for \( w - 2 \). In any state \( w < w \) for which \( \frac{dB_{w}}{du} < 0 \), \( \frac{dpw}{du} > 0 \) is determined by equation (2.8.11), \( \frac{dL_{w}}{du} < 0 \) is determined by (2.8.8), and \( \frac{dB_{w}}{du} < 0 \) is determined by (2.8.12). This completes the proof. \( \square \)

Together Lemmas 2.7 through 2.11 complete the proof of Propositions 2.6 and 2.7.

**Proof of Proposition 2.8**

Fixing the lender's behavior, the borrower's continuation utility in state \( n \) is

\[
B_n(p_n) = \frac{p_n (1 - e^{-\kappa dt}) + (1 - p_n) (1 - e^{-\gamma dt})}{1 - p_n e^{-(\rho + \kappa)dt} - (1 - p_n) e^{-(\rho + \gamma)dt}} e^{-\rho dt} \frac{u}{\rho}
\]

175
which increases linearly in $u$. Moving backward, suppose $B_{w+1}$ is increasing in $u$. Then, noting that

$$B_w(p_w) = \frac{p_w(1 - e^{-\kappa dt}) + (1 - p_w)(1 - e^{-\gamma dt})}{1 - p_w e^{-(\rho + \kappa)dt} - (1 - p_w) e^{-(\rho + \gamma)dt}} e^{-\rho dt} B_{w+1}$$

increases linearly in $B_{w+1}$ completes the proof.

**Proof of Proposition 2.9**

The proof for states $w \geq w$ proceeds exactly as in Proposition 2.6 and is thus omitted. In this section we provide an example in which $\frac{dp_{w+1}}{dp} < 0$ so that making the borrower more patient can strengthen the poverty trap.

We prove this result with a three state model $w \in \{1, 2, 3\}$ where the game ends if the borrower ever reaches state 3. We take

$$E = .15, q_1 = q_2 = q = 2, \phi = \frac{1}{2}, h = 100, T_1 = 600, T_2 = 1000, \frac{w}{\rho^B} = 2000 \text{ and } \rho^B = \rho^L = 1.$$ 

It is easily verified that Assumption 2.3 hold in states 1 and 2. That is,

$$\alpha^2 \rho^B = \left(\frac{3}{1.3}\right)^2 2000 > \frac{qE + h}{\rho^B} = 100.3$$

Now we verify that in state 2 the lender offers the borrower a restrictive contract with probability 1. If the borrower expects a restrictive contract with probability 1 then the lender gets the following continuation utility if he offers the borrower a restrictive contract in state 2.

$$L^R_2(1) = \frac{q(E + T_2) - T_2}{\rho^L} - \frac{u}{\rho^B} = 1000.3 - \frac{3}{1.3} 2000 \approx 539$$

If instead the lender offers the borrower an unrestricted contract at every period in state 2, his continuation utility is

$$L^U_2 = \frac{qT_2 - h - T_2}{\rho^L} (1 - \beta) \approx 9$$

Because the lender finds it *least* appealing to offer a restrictive contract when the borrower expects restrictive contracts with probability 1, we conclude that in the unique equilibrium the lender offers the borrower a restrictive contract with probability 1.

We next verify that in equilibrium, the lender mixes between restrictive and unrestricted contracts in state 1.

First, consider the lender’s continuation utility in state 1 from offering the borrower a restrictive
contract with probability 1 when she expects an restrictive contract with probability \( p \).

\[
L_1^R(p) = \frac{q(E + T_1) - T_1}{\rho^L} - \max \left\{ \frac{qE + h}{\rho^L}, \frac{\kappa}{\rho^B + \delta (p)} B_2 \right\}
\approx 600.3 - \max \left\{ 100.3, \left( \frac{.3}{1 + .3p + 100 (1 - p)} \right) \left( \frac{.3}{1.3} \right)^{2000} \right\}
\]

Note that the repayment rate the lender must set is the larger of \( qT - h \) and what the lender must set so that the borrower achieves the utility she would have received from investing \( E \) in fixed capital.

If instead the lender were to offer an unrestrictive contract with probability 1, her state 1 continuation utility would be

\[
L_1^U = \frac{qT - h - T_1}{\rho^L} (1 - \beta) + \beta L_2 \approx \frac{500}{101} + \frac{100}{101} 539
\]

It is easily verified that \( L_1^R(0) > L_1^U > L_1^R(1) \) and hence the unique equilibrium in state 1 involves the lender mixing between restrictive and unrestrictive contracts. The probability \( p_1 \) that the lender offers a restrictive contract is determined by the following equation.

\[
\frac{qT_1 - h - T_1}{\rho^L} (1 - \beta) + \beta L_2 = \frac{q(E + T_1) - T_1}{\rho^L} - \frac{\kappa}{\rho^B + \delta (p_1)} B_2
\Rightarrow
\frac{500}{101} + \frac{100}{101} \left( 1000.3 - \frac{.3}{1.3} 2000 \right) \approx 600.3 - \frac{.3}{1 + .3p_1 + 100 (1 - p_1)} \left( \frac{.3}{1.3} 2000 \right)
\Rightarrow
p_1 \approx .99
\]

Now, recall the investment rent in state 1 is

\[
\left( \beta - \frac{\kappa}{\rho^B + \delta (p_1)} \right) B_2 \approx \left( \frac{100}{101} - \frac{.3}{1 + .3p_1 + (1 - p_1) 100} \right) B_2
\]

We have

\[
d \left( \beta - \frac{\kappa}{\rho^B + \delta (p)} \right) B_2 = d \left( \beta - \frac{\kappa}{\rho^B + \delta (p)} \right) B_2 + \left( \beta - \frac{\kappa}{\rho^B + \delta (p)} \right) d \rho^B B_2
\]

177
Now,

\[
\frac{d}{dpB} B_2 = \frac{d}{dpB} \left( \frac{\rho B + 0.3}{\rho B + 0.3} \right) u 
\approx \frac{0.3}{(\rho B + 0.3)^2} \rho B - \frac{0.3}{\rho B + 0.3} \rho B 
\approx -\frac{0.3}{(1.3)^2} 2000 - \frac{0.3}{1.3} \cdot 2000 
\approx -816.56
\]

And,

\[
\frac{d}{dpB} \left( \frac{100}{\rho B + 100} - \frac{0.3}{\rho B + 0.3 + (1 - p) 100} \right) 
\approx \left( \frac{100}{(\rho B + 100)^2} + \frac{0.3}{(\rho B + 0.3 + (1 - p) 100)^2} \right) 
\approx .05
\]

So,

\[
\frac{d}{dpB} \left[ (\beta - \frac{\kappa}{\rho B + \delta}) B_2 \right] \approx -675.87 < 0
\]

Therefore making the borrower more patient increases the investment rent and reduces \( p_1 \).

**Proof of Proposition 2.10**

Fix a game \( \Gamma \) with \( n \) states, and a cost of fixed investment \( \{\phi_w\} \). Then for game \( \Gamma^m \) with \( m > 0 \), a borrower investing in fixed capital rate \( i \) in state \( 2mn \) will derive a state \( 2mn \) continuation value of

\[
\frac{i}{\phi_{2mn}} \frac{u}{\rho} + \frac{1}{\phi_{2mn}} \frac{u}{\rho}
\]

which converges to \( \frac{u}{\rho} \) as \( m \) becomes large. If the borrower's equilibrium expectation is that the lender will offer the restrictive contract with probability 1, then the lender's continuation utility in state \( 2mn \) from doing so is

\[
L^R_{2mn}(1) = \frac{q_{2mn}(E_{2mn} + T_{2mn}) - T_{2mn}}{\rho} - \frac{\kappa_{2mn} u}{\rho + \kappa_{2mn}}
\]

which for sufficiently large \( m \) will be negative when it is socially efficient to invest.

On the other hand the lender's continuation utility in state \( n \) if he offers an unrestrictive contract in every period is

\[
L^U_{2mn} = \frac{q_{2mn}T_{2mn} - h - T_{2mn}}{\rho} \left( 1 - \frac{\gamma_{2mn}}{\rho + \gamma_{2mn}} \right)
\]

178
which is positive for all $m > 0$. Thus for sufficiently high $m$, the lender will offer an unrestrictive contract with positive probability in state $2^m n$, completing the proof.

**Proof of Proposition 2.11**

For states $w > \bar{w}_{m-1}$, the proof closely follows that of Proposition 2.6. Specifically, for states $w > \bar{w}_m$, the proof follows that of Lemma 2.7. For states $w \in \{\bar{w}_m, \ldots, \bar{w}_m\}$ the proof follows that of Lemma 2.8. And for states $w \in \{\bar{w}_{m-1} + 1, \ldots, \bar{w}_m - 1\}$ the proof follows that of Lemma 2.10.

For $w \in \{\bar{w}_{m-2} + 1, \bar{w}_{m-1}\}$ the logic of Proposition 2.6 is reversed, as $\frac{dB_{\bar{w}_{m-1}+1}}{du} < 0$. That is, in the state directly beyond the pure restrictive state $\bar{w}_{m-1}$, the borrower’s continuation utility is decreasing in $u$. In contrast, in the state directly beyond the pure restrictive state $\bar{w}_m$, the borrower’s continuation utility is increasing in $u$. So, the comparative static for $w \in \{\bar{w}_{m-2} + 1, \bar{w}_{m-1}\}$ comes directly from reversing the signs in Lemmas 2.8 and 2.10.

The proof proceeds similarly for all consumption regions $\{\bar{w}_m, \ldots, \bar{w}_\bar{m}\}$ for $\bar{m} \leq m - 2$.

**Proof of Proposition 2.12**

**Lemma 2.12.** In any equilibrium in which the borrower never accepts the entrant’s contract, the borrower must always be weakly better off than she would be from receiving the contract $(T_w, I)$ or $(T_w, C)$ from the entrant in state $w$.

*Proof.* Consider any equilibrium in which the borrower never accepts the entrant’s contract. In such an equilibrium the entrant’s continuation value in any state is 0. Suppose toward contradiction that there is some state $w$ in which with positive probability the incumbent offers the borrower a contract $\tilde{c}$ that provides her with strictly less utility than she would receive from accepting the contract $(T_w, I)$ or $(T_w, C)$ from the entrant. Then there exists an $\epsilon > 0$ such that the borrower would strictly prefer the entrant’s contract $(T_w + \epsilon, I)$ or $(T_w + \epsilon, C)$ to the incumbent’s contract $\tilde{c}$. Therefore by deviating to offer one of these contracts, the entrant could guarantee himself a positive payoff, contradicting the premise that the borrower never accepts the entrant’s contract in equilibrium.

Lemma 2.12 implies that in equilibrium the borrower’s outside option is the maximum of the utility she would receive from accepting the entrant’s contract $(T_w, I)$ or $(T_w, C)$, and the utility she would receive from allocating her autarkic endowment flexibly.

The remainder of the proof proceeds exactly as in Proposition 2.2, with this new individual rationality constraint in place of the borrower’s autarkic constraint, and is thus omitted.
Proof of Proposition 2.13

We first identify a $\tilde{\psi}$ such that for $\psi > \tilde{\psi}$, the equilibrium is the same as in the monopolist case. We determine $\tilde{\psi}$ as follows. Let $B^\text{max}_{w+1}$ be the maximal feasible (potentially out of equilibrium) state $w+1$ continuation utility that the borrower can achieve. Further suppose that in state $w$ the borrower anticipates the incumbent will never make a loan offer, inducing her to have maximal feasible value of investment in state $w$. Then if she opts not to borrow from the entrant, her state $w$ continuation utility is

$$B_w = e^{-\rho dt} \left( (1 - e^{-r dt}) B^\text{max}_{w+1} + e^{-\kappa dt} B_w \right)$$

Let $\tilde{\psi}_w$ satisfy

$$e^{-\rho dt} \left( 1 - e^{-\frac{e_{w+1} - \kappa_w}{\psi_w} dt} \right) B^\text{max}_{w+1} + e^{-\frac{e_w + \tau_w - \kappa_w}{\psi_w} dt} B_w - \tilde{\psi}_w dt < e^{-\rho dt} \left( (1 - e^{-r dt}) B^\text{max}_{w+1} + e^{-\kappa dt} B_w \right)$$

ensuring that the borrower would prefer to invest her autarkic endowment rather than borrow from the entrant even when her value from investment is as high as is feasible in state $w$. Now let $\tilde{\psi} \equiv \max_w \tilde{\psi}_w$. Clearly in equilibrium, if the borrower ever rejects the incumbent’s contract then she will choose not to borrow from the entrant and her individual rationality constraint will be the same as in Proposition 2.2. Thus the equilibrium outcome will be the same as in the monopolist case.

We now prove the existence of a $\psi > 0$ such that for $\psi < \tilde{\psi}$ the incumbent offers an unrestricted contract with probability 1 in every period. This is so because by Lemma 2.12, for $\psi = 0$ the borrower’s outside option is to take whichever she prefers of the maximally generous restrictive and unrestricted contracts from the entrant at no additional cost. So long as it is efficient to invest in business expansion, the borrower will always prefer the maximally unrestricted contract. By continuity, there exists a $\psi > 0$ such that the same will be true for any $\psi < \tilde{\psi}$. So for $\psi < \tilde{\psi}$ the incumbent must offer unrestricted contracts in equilibrium. And by the same logic, as $\psi \to 0$, the incumbent’s equilibrium contract offer must converge to the maximally generous unrestricted contract. This completes the proof.

Proof of Proposition 2.14

Lemma 2.13. If business growth is efficient and $p_n \in (0, 1)$, $\frac{dp_n}{d\psi} \geq 0$, $\frac{dp_n}{d\psi} \leq 0$, and $\frac{dp_n}{d\psi} = 0$.

Proof. Suppose that in state $n$ the equilibrium probability the incumbent lender offers a restrictive contract is $p_n \in (0, 1)$. If $\psi$ is sufficiently high that the borrower’s outside option is to invest her own autarkic endowment rather than borrow from the entrant, then $\frac{dp_n}{d\psi} = 0$ and the conclusion of
the lemma is satisfied. Else the borrower’s outside option is to borrow from the entrant and incur the non pecuniary cost of $\psi dt$.

Now suppose that in equilibrium the borrower derives more utility from the maximally extractive unrestricted contract than she does from the maximally extractive but acceptable restrictive contract (and hence her expansion rent is positive). Then $p_n$ will be determined as it was in Lemma 2.7, respecting the borrower’s modified individual rationality constraint. Decreasing $\psi$ has the same impact as increasing $U$, it improves the borrower’s outside option so that she becomes more demanding of restrictive contracts. As in Lemma 2.7 this causes an equilibrium shift towards unrestricted contracts, so $\frac{dp_n}{d\psi_w} > 0$. The lender remains indifferent between the two types of contracts so his continuation utility $L_n$ is unaffected.

Last suppose that in equilibrium the borrower derives the same utility from the maximally extractive unrestricted contract as she does from the maximally extractive but acceptable restrictive contract (and hence her expansion rent is zero). Then by the logic is Proposition 2.5, it must be that $p_n = 0$, contradicting the premise of the lemma.

We complete the proof of the proposition by backward induction.

**Lemma 2.14.** Consider state $w$ for which $p_w \in (0, 1)$ for and for which $\frac{dL_{w+1}}{d\psi} = 0$ and $\frac{dB_{w+1}}{d\psi} \leq 0$. Then $\frac{dp_w}{d\psi} \geq 0$.

*Proof.* The proof for this lemma proceeds in the same way as the proof for Lemma 2.13, however when considering how demanding the borrower is of restrictive contracts, her outside option now improves both due to the increase in competition in the current state and to her improved continuation value in state $w + 1$. Otherwise the proof is the same and is thus omitted.

Together Lemmas 2.13 and 2.14 complete the proof of the result.

**Proof of Proposition 2.15**

**Lemma 2.15.** The borrower never rejects an unrestricted contract on the equilibrium path.

*Proof.* Assumption 2.6 guarantees that if the borrower receives an unrestricted contract, she necessarily invests more in the fixed capital project than she could have in autarky. Thus by the logic in Lemma 2.3, the borrower would only ever reject the maximally extractive unrestricted contract if in equilibrium she gets a more generous unrestricted contract with certainty.

**Lemma 2.16.** The borrower may reject a restrictive contract on the equilibrium path.

*Proof.* To prove this lemma we need only find an example in which the borrower rejects a restrictive contract with positive probability. To do so we modify the example from the proof of Proposition
2.9. Specifically consider the one state example in which we take $E = .15$, $q = 2$, $\phi = \frac{1}{2}$, $h = 100$, $T = 1000$, $\frac{\mu}{\sigma^2} = 2000$ and $\rho^R = \rho^L = 1$. We define the distribution $G$ such that $\nu = 0$ with probability $1 - \varepsilon$, and $\nu = 45$ with probability $\varepsilon$.

We verified in the proof of Proposition 2.9 that this example satisfies Assumption 2.3 and that in equilibrium the lender offers the borrower a restrictive contract with probability 1. Clearly for sufficiently small $\varepsilon$, the lender would prefer to offer the least generous restrictive contract that borrowers of type $\nu = 0$ would accept. The loss the lender suffers from being rejected with probability $\varepsilon$ is vanishing. In contrast, if the lender offers a contract that both types of borrowers would accept, he incurs a first order loss in order to compensate the high type borrower for the $\nu = 45$ additional forgone investment.

Proof of Proposition 2.16

Define $\beta_s \equiv \gamma^\frac{\varepsilon}{\rho + \gamma - \frac{T}{E}}$. Now suppose in state $w$ the borrower anticipates a restrictive contract with probability 1. It is straightforward to show that the borrower's expansion rent when she can invest a fraction of her endowment $s$ flexibly no matter the contractual restriction is $(\beta_s - \alpha) B_{w+1}$. Further as in Proposition 2.4, in equilibrium the borrower receives a restrictive contract with certainty in state $w$ if and only if

$$(\beta_s - \alpha) B_{w+1} \geq \beta \left( B_{w+1} + L_{w+1} - \frac{q_w (E + T_w) - T_w}{\rho} - \phi \right).$$

For $s < \frac{E_w}{T_w + L_w}$, the borrower will grow more slowly in equilibrium than in autarky in any state in which the above condition is satisfied.

2.8.3 Tables and Figures
Table 2.8.1: Default is U-Shaped in Income in Field et. al. Data

<table>
<thead>
<tr>
<th></th>
<th>(1) Doesn't Complete Repayment</th>
<th>(2) Doesn't Complete Repayment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Profits</td>
<td>-0.153</td>
<td>-0.167*</td>
</tr>
<tr>
<td></td>
<td>(0.0955)</td>
<td>(0.0945)</td>
</tr>
<tr>
<td>Log Profits Sq.</td>
<td>0.0108</td>
<td>0.0118*</td>
</tr>
<tr>
<td></td>
<td>(0.00684)</td>
<td>(0.00672)</td>
</tr>
<tr>
<td>N</td>
<td>660</td>
<td>660</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors in parentheses. All columns are borrower level OLS regressions. The outcome variable is whether the borrower has completed repayment a year after disbursement. Controls are those included in Field et. al. (2013) and include borrower education, household size, religion, literacy, marital status, age, household shocks, business ownership at baseline, financial control, home ownership, and whether the household has a drain. Log Profits are measured three years after loan disbursement - this is the publicly available measure.

* p < 0.10, ** p < 0.05, *** p < 0.01
Table 2.8.2: Village Covariates Correlated With Inverse Size

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Village Head</td>
<td>Village Head</td>
<td>Number</td>
<td>Fraction</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>Education</td>
<td>Agricultural</td>
<td>with Electricity</td>
</tr>
<tr>
<td>Inv. Size</td>
<td>245.0</td>
<td>-532.3**</td>
<td>-115.2**</td>
<td>10.32</td>
</tr>
<tr>
<td></td>
<td>(248.7)</td>
<td>(210.9)</td>
<td>(51.99)</td>
<td>(7.317)</td>
</tr>
<tr>
<td>Constant</td>
<td>44.44***</td>
<td>20.37***</td>
<td>4.079***</td>
<td>0.778***</td>
</tr>
<tr>
<td></td>
<td>(3.041)</td>
<td>(2.909)</td>
<td>(0.659)</td>
<td>(0.112)</td>
</tr>
<tr>
<td>N</td>
<td>55</td>
<td>55</td>
<td>55</td>
<td>55</td>
</tr>
</tbody>
</table>

|                      | (5)          | (6)          | (7)          | (8)          |
|                      | Fraction     | Fraction     | Fraction     | Fraction     |
|                      | with TV      | with Phone   | with Tractor | with Chemical |
| Inv. Size            | -63.27       | -67.78       | 1.515        | 7.406        |
|                      | (62.63)      | (64.54)      | (8.970)      | (8.377)      |
| Constant             | 1.704*       | 1.032        | 0.273***     | 0.721***     |
|                      | (0.933)      | (0.964)      | (0.0967)     | (0.114)      |
| N                    | 55           | 55           | 54           | 55           |

|                      | (9)          | (10)         | (11)         | (12)         |
|                      | Has Factory  | Has Temple   | Distance to  | Has Common   |
|                      |              |              | Main Road    | Land         |
| Inv. Size            | -25.92       | -10.59       | -57.89       | -47.89*      |
|                      | (26.46)      | (31.76)      | (193.9)      | (24.91)      |
| Constant             | 1.772***     | 1.762***     | 7.478*       | 2.068***     |
|                      | (0.312)      | (0.351)      | (3.949)      | (0.316)      |
| N                    | 55           | 55           | 55           | 55           |

|                      | (13)         | (14)         | (15)         | (16)         |
|                      | Soil Problems| Fraction     | Irrigation   | Fraction     |
|                      |              | Agricultural |             | Wage Workers |
| Inv. Size            | -1.867       | 11.45        | 6.571        | 24.43        |
|                      | (33.51)      | (7.393)      | (21.76)      | (17.26)      |
| Constant             | 1.819***     | 0.663***     | 2.642***     | 0.289        |
|                      | (0.366)      | (0.111)      | (0.249)      | (0.199)      |
| N                    | 55           | 54           | 55           | 55           |

Notes: Robust standard errors in parentheses. All columns are village level OLS regressions.
* p < 0.10, ** p < 0.05, *** p < 0.01
Table 2.8.3: Restrictive Loans are Correlated with Lower Income Expectation

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Expected Income</td>
<td>Log Expected Income</td>
<td>Log Expected Income</td>
<td>Log Expected Income</td>
</tr>
<tr>
<td>Restrictive</td>
<td>-0.110**</td>
<td>-0.193***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0531)</td>
<td>(0.0528)</td>
<td></td>
</tr>
<tr>
<td>Log Income</td>
<td>0.693***</td>
<td>0.612***</td>
<td>0.688***</td>
</tr>
<tr>
<td></td>
<td>(0.0422)</td>
<td>(0.0651)</td>
<td>(0.0422)</td>
</tr>
<tr>
<td>Log Loan Size</td>
<td>-0.00210</td>
<td>-0.00310</td>
<td>0.00559</td>
</tr>
<tr>
<td></td>
<td>(0.0187)</td>
<td>(0.0177)</td>
<td>(0.0180)</td>
</tr>
<tr>
<td>Land - borrower use</td>
<td>-0.0401</td>
<td>-0.103</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0674)</td>
<td>(0.0809)</td>
<td></td>
</tr>
<tr>
<td>Land - lender use</td>
<td>-0.275**</td>
<td>-0.345***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.132)</td>
<td>(0.120)</td>
<td></td>
</tr>
<tr>
<td>Future Crop</td>
<td>0.0644</td>
<td>-0.0893</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.182)</td>
<td>(0.127)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>-0.238</td>
<td>-0.199**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.163)</td>
<td>(0.0935)</td>
<td></td>
</tr>
<tr>
<td>Single Guarantor</td>
<td>0.0705</td>
<td>-0.0919</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0609)</td>
<td>(0.0651)</td>
<td></td>
</tr>
<tr>
<td>Mult. Guarantor</td>
<td>-0.182**</td>
<td>-0.191**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0753)</td>
<td>(0.0743)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>1999</th>
<th>1783</th>
<th>1999</th>
<th>1783</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave FE s</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Village FE s</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Household FE s</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors clustered at the village level in parentheses. All columns are loan level OLS regressions. The outcome variable is the log of expected income in the next year. Land - borrower use is an indicator taking the value 1 if land is used as collateral but the borrower uses it. Land - lender use is an indicator taking the value of 1 if the borrower forfeits his land to the lender. Income and loan size are trimmed at the 99th percentile.

* p < 0.10, ** p < 0.05, *** p < 0.01
### Table 2.8.4: Restrictive Loans Are Correlated with Lower Interest Rate

<table>
<thead>
<tr>
<th></th>
<th>(1) Monthly Interest</th>
<th>(2) Monthly Interest</th>
<th>(3) Monthly Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restrictive</td>
<td>-0.0328***</td>
<td>-0.0326***</td>
<td>-0.0168</td>
</tr>
<tr>
<td></td>
<td>(0.00910)</td>
<td>(0.0101)</td>
<td>(0.0125)</td>
</tr>
<tr>
<td>Log Income</td>
<td>-0.00423</td>
<td>0.000321</td>
<td>0.000263</td>
</tr>
<tr>
<td></td>
<td>(0.00384)</td>
<td>(0.00515)</td>
<td>(0.00737)</td>
</tr>
<tr>
<td>N</td>
<td>1101</td>
<td>1068</td>
<td>878</td>
</tr>
<tr>
<td>Wave FEs</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Village FEs</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Household FEs</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors clustered at the village level in parentheses. All columns are loan level OLS regressions. The outcome variable is the monthly interest rate. We do not control for loan size as it is used to construct interest rates. Income and interest rate are trimmed at the 99th percentile.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

### Table 2.8.5: Main Comparative Static on Village Fund Intensity

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Restrictive</td>
<td>Restrictive</td>
<td>Restrictive</td>
</tr>
<tr>
<td>Inv. Size</td>
<td>8.893*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.126)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>0.222***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0815)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post*Inv. Size</td>
<td>-10.87*</td>
<td>-13.27**</td>
<td>-21.65***</td>
</tr>
<tr>
<td></td>
<td>(6.364)</td>
<td>(5.351)</td>
<td>(5.572)</td>
</tr>
<tr>
<td>Log Loan Size</td>
<td>0.0613***</td>
<td>0.0421***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0142)</td>
<td>(0.0140)</td>
<td></td>
</tr>
<tr>
<td>Log Income</td>
<td>-0.00516</td>
<td>-0.00320</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0126)</td>
<td>(0.00957)</td>
<td></td>
</tr>
<tr>
<td>Village Fund</td>
<td>0.0770</td>
<td>0.130***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0581)</td>
<td>(0.0296)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1710</td>
<td>1710</td>
<td>1620</td>
</tr>
<tr>
<td>Wave FEs</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Village FEs</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Household FEs</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors clustered at the village level in parentheses. All columns are loan level OLS regressions. The outcome variable is whether the money lender demands a restrictive form of collateral. Controls include number of agricultural cooperatives, whether the village has common land, and the education of the village head as well as their interaction with post. Income and loan size are trimmed at the 99th percentile. Only villages with between 50 and 250 households are included.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
### Table 2.8.6: Parallel Trends Prior to 2002

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Restrictive</td>
<td>Restrictive</td>
<td>Restrictive</td>
</tr>
<tr>
<td>Wave</td>
<td>0.00787</td>
<td>0.0115</td>
<td>-0.0183</td>
</tr>
<tr>
<td></td>
<td>(0.0321)</td>
<td>(0.0298)</td>
<td>(0.0183)</td>
</tr>
<tr>
<td>Inv. Size</td>
<td>-625.4</td>
<td>-0.121</td>
<td>2.841</td>
</tr>
<tr>
<td></td>
<td>(5916.8)</td>
<td>(3.086)</td>
<td>(1.709)</td>
</tr>
<tr>
<td>Wave*Inv. Size</td>
<td>0.317</td>
<td>-0.121</td>
<td>2.841</td>
</tr>
<tr>
<td></td>
<td>(2.960)</td>
<td>(3.086)</td>
<td>(1.709)</td>
</tr>
<tr>
<td>Log Loan Size</td>
<td>0.0453***</td>
<td>0.0146</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0159)</td>
<td>(0.0157)</td>
<td></td>
</tr>
<tr>
<td>Log Income</td>
<td>-0.0282*</td>
<td>-0.0233</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0150)</td>
<td>(0.0179)</td>
<td></td>
</tr>
<tr>
<td>Village Fund</td>
<td>-0.00676</td>
<td>0.107</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0411)</td>
<td>(0.0734)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>916</td>
<td>914</td>
<td>826</td>
</tr>
<tr>
<td>Household FEs</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Village FEs</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors clustered at the village level in parentheses. All columns are loan level OLS regressions. The outcome variable is whether the money lender demands a restrictive form of collateral. Controls include number of agricultural cooperatives, whether the village has common land, and the education of the village head as well as their interaction with post. Income and loan size are trimmed at the 99th percentile. Only villages with between 50 and 250 households are included.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 2.8.7: Whether a Household Borrows from Money Lenders Regressed on \( post*invsize \)

<table>
<thead>
<tr>
<th></th>
<th>(1) Borrow From Money Lender</th>
<th>(2) Borrow From Money Lender</th>
<th>(3) Borrow From Money Lender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inv. Size</td>
<td>3.421</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.093)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>-0.109***</td>
<td>-0.101</td>
<td>-0.0637</td>
</tr>
<tr>
<td></td>
<td>(0.0378)</td>
<td>(0.0729)</td>
<td>(0.0788)</td>
</tr>
<tr>
<td>Post*Inv. Size</td>
<td>1.810</td>
<td>-0.739</td>
<td>-0.528</td>
</tr>
<tr>
<td></td>
<td>(3.427)</td>
<td>(3.471)</td>
<td>(3.798)</td>
</tr>
<tr>
<td>Log Income</td>
<td>-0.0127</td>
<td>-0.00263</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00954)</td>
<td>(0.00848)</td>
<td></td>
</tr>
<tr>
<td>Village Fund</td>
<td>0.0238</td>
<td>-0.0251</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0156)</td>
<td>(0.0196)</td>
<td></td>
</tr>
<tr>
<td>( N )</td>
<td>6782</td>
<td>6782</td>
<td>6715</td>
</tr>
<tr>
<td>Wave FEs</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Village FEs</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Household FEs</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors clustered at the village level in parentheses. All columns are villager x time level OLS regressions. The outcome variable is whether the villager borrows from a money lender in a given survey wave. Controls include number of agricultural cooperatives, whether the village has common land, and the education of the village head as well as their interaction with post. Income and loan size are trimmed at the 99th percentile. Only villages with between 50 and 250 households are included.

* \( p < 0.10 \), ** \( p < 0.05 \), *** \( p < 0.01 \)
Table 2.8.8: Total Household Borrowing from Money Lenders Regressed on post*invsize

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Log Total</td>
<td>Log Total</td>
<td>Log Total</td>
</tr>
<tr>
<td></td>
<td>Borrowing From</td>
<td>Borrowing From</td>
<td>Borrowing From</td>
</tr>
<tr>
<td></td>
<td>Money Lenders</td>
<td>Money Lenders</td>
<td>Money Lenders</td>
</tr>
<tr>
<td>Inv. Size</td>
<td>39.29</td>
<td>(60.64)</td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>-0.964**</td>
<td>-0.920</td>
<td>-0.586</td>
</tr>
<tr>
<td></td>
<td>(0.368)</td>
<td>(0.688)</td>
<td>(0.771)</td>
</tr>
<tr>
<td>Post*Inv. Size</td>
<td>13.60</td>
<td>-11.82</td>
<td>-5.858</td>
</tr>
<tr>
<td></td>
<td>(33.72)</td>
<td>(34.36)</td>
<td>(38.65)</td>
</tr>
<tr>
<td>Log Income</td>
<td>-0.0505</td>
<td>0.00934</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0956)</td>
<td>(0.0896)</td>
<td></td>
</tr>
<tr>
<td>Village Fund</td>
<td>0.308*</td>
<td>-0.216</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.159)</td>
<td>(0.195)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>6782</td>
<td>6782</td>
<td>6715</td>
</tr>
<tr>
<td>Wave FEs</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Village FEs</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Household FEs</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors clustered at the village level in parentheses. All columns are villager x time level OLS regressions. The outcome variable is the log of how much the villager borrows from a money lender in a given survey wave. Controls include number of agricultural cooperatives, whether the village has common land, and the education of the village head as well as their interaction with post. Income and loan size are trimmed at the 99th percentile. Only villages with between 50 and 250 households are included.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 2.8.9: Heckman Selection Model

<table>
<thead>
<tr>
<th></th>
<th>(1) Restrictive</th>
<th>(2) Restrictive</th>
<th>(3) Restrictive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inv. Size</td>
<td>8.803*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.206)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>0.231***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0896)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post*Inv. Size</td>
<td>-10.96*</td>
<td>-13.10**</td>
<td>-37.79</td>
</tr>
<tr>
<td></td>
<td>(6.355)</td>
<td>(5.353)</td>
<td>(70.73)</td>
</tr>
<tr>
<td>Log Loan Size</td>
<td>0.0613***</td>
<td>0.0414</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0141)</td>
<td>(0.0587)</td>
<td></td>
</tr>
<tr>
<td>Log Income</td>
<td>-0.000382</td>
<td>-0.451</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0136)</td>
<td>(1.661)</td>
<td></td>
</tr>
<tr>
<td>Village Fund</td>
<td>0.0771</td>
<td>0.124</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0578)</td>
<td>(0.214)</td>
<td></td>
</tr>
</tbody>
</table>

Selection Equation

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inv. Size</td>
<td>0.919</td>
<td>0.837</td>
<td>0.872</td>
</tr>
<tr>
<td></td>
<td>(16.59)</td>
<td>(16.60)</td>
<td>(4.937)</td>
</tr>
<tr>
<td>Post</td>
<td>-0.244</td>
<td>-0.247</td>
<td>-0.247**</td>
</tr>
<tr>
<td></td>
<td>(0.172)</td>
<td>(0.171)</td>
<td>(0.120)</td>
</tr>
<tr>
<td>Post*Inv. Size</td>
<td>-5.092</td>
<td>-5.015</td>
<td>-5.018</td>
</tr>
<tr>
<td></td>
<td>(9.189)</td>
<td>(9.184)</td>
<td>(6.610)</td>
</tr>
<tr>
<td>Log Income</td>
<td>-0.134***</td>
<td>-0.135***</td>
<td>-0.135***</td>
</tr>
<tr>
<td></td>
<td>(0.0310)</td>
<td>(0.0310)</td>
<td>(0.0136)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>24916</td>
<td>24916</td>
<td>24916</td>
</tr>
<tr>
<td>Wave FEs</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Village FEs</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Household FEs</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors clustered at the village level in parentheses. All columns are loan level Heckman Selection models under the assumption that errors are jointly normally distributed. The outcome variable is whether the money lender demands a restrictive form of collateral. Controls include number of agricultural cooperatives, whether the village has common land, and the education of the village head as well as their interaction with post. Controls in the selection equation include all non fixed effect regressors except for loan size in main specification and whether the household's primary means of income generation is farm or nonfarm work. Income and loan size are trimmed at the 99th percentile. Only villages with between 50 and 250 households are included.

* p < 0.10, ** p < 0.05, *** p < 0.01
Table 2.8.10: Placebo Test: Main Regression with Population of Borrowers From Neighbors Rather Than Money Lenders

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Restrictive</td>
<td>Restrictive</td>
<td>Restrictive</td>
</tr>
<tr>
<td>Inv. Size</td>
<td>-2.981</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.829)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>0.00409</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.147)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post*Inv. Size</td>
<td>7.824</td>
<td>18.80</td>
<td>-6.507</td>
</tr>
<tr>
<td></td>
<td>(12.56)</td>
<td>(11.48)</td>
<td>(12.28)</td>
</tr>
<tr>
<td>Log Loan Size</td>
<td>0.0971***</td>
<td>0.0890***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0207)</td>
<td>(0.0228)</td>
<td>(0.0228)</td>
</tr>
<tr>
<td>Log Income</td>
<td>-0.0467*</td>
<td>-0.0151</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0268)</td>
<td>(0.0230)</td>
<td>(0.0230)</td>
</tr>
<tr>
<td>Village Fund</td>
<td>0.0439</td>
<td>-0.0869</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0693)</td>
<td>(0.0677)</td>
<td>(0.0677)</td>
</tr>
<tr>
<td>N</td>
<td>746</td>
<td>745</td>
<td>649</td>
</tr>
<tr>
<td>Wave FEs</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Village FEs</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Household FEs</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors clustered at the village level in parentheses. All columns are loan level OLS regressions. The outcome variable is whether the neighbor demands a restrictive form of collateral. Controls include number of agricultural cooperatives, whether the village has common land, and the education of the village head as well as their interaction with post. Income and loan size are trimmed at the 99th percentile. Only villages with between 50 and 250 households are included.

* p < 0.10, ** p < 0.05, *** p < 0.01
Table 2.8.11: Placebo Test: Main Regression with Unrestrictive Collateral as Outcome Variable

<table>
<thead>
<tr>
<th></th>
<th>(1) Unrestrictive</th>
<th>(2) Unrestrictive</th>
<th>(3) Unrestrictive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inv. Size</td>
<td>1.503</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.379)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>-0.0971</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0725)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post*Inv. Size</td>
<td>4.165</td>
<td>12.03**</td>
<td>14.34***</td>
</tr>
<tr>
<td></td>
<td>(5.664)</td>
<td>(5.869)</td>
<td>(5.366)</td>
</tr>
<tr>
<td>Log Loan Size</td>
<td>0.0878***</td>
<td>0.0665***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0147)</td>
<td>(0.0155)</td>
<td></td>
</tr>
<tr>
<td>Log Income</td>
<td>-0.0331**</td>
<td>-0.00913</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0144)</td>
<td>(0.0135)</td>
<td></td>
</tr>
<tr>
<td>Village Fund</td>
<td>-0.128***</td>
<td>-0.0177</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0435)</td>
<td>(0.0421)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1710</td>
<td>1710</td>
<td>1620</td>
</tr>
<tr>
<td>Wave FEs</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Village FEs</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Household FEs</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors clustered at the village level in parentheses. All columns are loan level OLS regressions. The outcome variable is whether the money lender demands an unrestrictive form of collateral. Controls include number of agricultural cooperatives, whether the village has common land, and the education of the village head as well as their interaction with post. Income and loan size are trimmed at the 99th percentile. Only villages with between 50 and 250 households are included.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 2.8.12: Main Regression with Interest Rates as Outcome Variable

<table>
<thead>
<tr>
<th></th>
<th>(1) Monthly Interest</th>
<th>(2) Monthly Interest</th>
<th>(3) Monthly Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inv. Size</td>
<td>-0.547</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.186)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>-0.0456**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0195)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post*Inv. Size</td>
<td>1.959</td>
<td>0.512</td>
<td>-0.197</td>
</tr>
<tr>
<td></td>
<td>(1.751)</td>
<td>(2.104)</td>
<td>(2.056)</td>
</tr>
<tr>
<td>Log Income</td>
<td>-0.00366</td>
<td>-0.00254</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00455)</td>
<td>(0.00697)</td>
<td></td>
</tr>
<tr>
<td>Village Fund</td>
<td>-0.0362*</td>
<td>-0.0395***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0208)</td>
<td>(0.0140)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>860</td>
<td>859</td>
<td>757</td>
</tr>
<tr>
<td>Wave FEs</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Village FEs</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Household FEs</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors clustered at the village level in parentheses. All columns are loan level OLS regressions. The outcome variable is the monthly interest rate. Controls include number of agricultural cooperatives, whether the village has common land, and the education of the village head as well as their interaction with post, but not loan size as it is used to construct the outcome variable. Income and interest rate are trimmed at the 99th percentile. Only villages with between 50 and 250 households are included.

* _p < 0.10_, ** _p < 0.05_, *** _p < 0.01_
Table 2.8.13: Robustness Check: Main Comparative Static Without Restriction on Village Size

<table>
<thead>
<tr>
<th></th>
<th>(1) Restrictive</th>
<th>(2) Restrictive</th>
<th>(3) Restrictive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inv. Size</td>
<td>6.259**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.089)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>0.205***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0570)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.119)</td>
<td>(3.969)</td>
<td>(5.988)</td>
</tr>
<tr>
<td>Log Loan Size</td>
<td>0.0628***</td>
<td>0.0408***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0137)</td>
<td>(0.0137)</td>
<td></td>
</tr>
<tr>
<td>Log Income</td>
<td>-0.00170</td>
<td>-0.00207</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0122)</td>
<td>(0.00916)</td>
<td></td>
</tr>
<tr>
<td>Village Fund</td>
<td>0.0821</td>
<td>0.125***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0558)</td>
<td>(0.0290)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1826</td>
<td>1824</td>
<td>1721</td>
</tr>
<tr>
<td>Wave FEs</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Village FEs</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Household FEs</td>
<td>No</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors clustered at the village level in parentheses. All columns are loan level OLS regressions. The outcome variable is whether the money lender demands a restrictive form of collateral. Controls include number of agricultural cooperatives, whether the village has common land, and the education of the village head as well as their interaction with post. Income and loan size are trimmed at the 99th percentile.

* p < 0.10, ** p < 0.05, *** p < 0.01
Table 2.8.14: Robustness Check: Main Comparative Static Using Data Only From 1999-2004

<table>
<thead>
<tr>
<th></th>
<th>(1) Restrictive</th>
<th>(2) Restrictive</th>
<th>(3) Restrictive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inv. Size</td>
<td>8.374</td>
<td>5.995</td>
<td>0.0842</td>
</tr>
<tr>
<td>Post</td>
<td>0.196**</td>
<td>0.0842</td>
<td>(0.0842)</td>
</tr>
<tr>
<td>Post*Inv. Size</td>
<td>-10.35</td>
<td>-7.474</td>
<td>-11.51**</td>
</tr>
<tr>
<td>Log Loan Size</td>
<td>0.0609***</td>
<td>0.0543***</td>
<td>0.0160</td>
</tr>
<tr>
<td>Log Income</td>
<td>-0.00407</td>
<td>-0.00658</td>
<td>0.0176</td>
</tr>
<tr>
<td>Village Fund</td>
<td>0.0635</td>
<td>0.0888***</td>
<td>0.0646</td>
</tr>
<tr>
<td>N</td>
<td>1189</td>
<td>1188</td>
<td>1124</td>
</tr>
<tr>
<td>Wave FEs</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Village FEs</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Household FEs</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors clustered at the village level in parentheses. All columns are loan level OLS regressions. The outcome variable is whether the money lender demands a restrictive form of collateral. Controls include number of agricultural cooperatives, whether the village has common land, and the education of the village head as well as their interaction with post. Income and loan size are trimmed at the 99th percentile. Only villages with between 50 and 250 households and only waves collected between 1999 and 2004 are included.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 2.8.15: Robustness Check: Main Comparative Static with Data Driven Definition of Restrictiveness

<table>
<thead>
<tr>
<th></th>
<th>(1) Restrictive</th>
<th>(2) Restrictive</th>
<th>(3) Restrictive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inv. Size</td>
<td>0.0578</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.149)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>0.186**</td>
<td>0.224</td>
<td>0.320**</td>
</tr>
<tr>
<td></td>
<td>(0.0811)</td>
<td>(0.137)</td>
<td>(0.138)</td>
</tr>
<tr>
<td>Post*Inv. Size</td>
<td>-8.709</td>
<td>-7.608</td>
<td>-15.55***</td>
</tr>
<tr>
<td></td>
<td>(6.490)</td>
<td>(5.609)</td>
<td>(5.553)</td>
</tr>
<tr>
<td>Log Loan Size</td>
<td>0.0444***</td>
<td>0.0148</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0119)</td>
<td>(0.0115)</td>
<td></td>
</tr>
<tr>
<td>Log Income</td>
<td>-0.00607</td>
<td>-0.00491</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0112)</td>
<td>(0.00949)</td>
<td></td>
</tr>
<tr>
<td>Village Fund</td>
<td>0.0597*</td>
<td>0.0842***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0326)</td>
<td>(0.0290)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1710</td>
<td>1710</td>
<td>1620</td>
</tr>
<tr>
<td>Wave FEs</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Village FEs</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Household FEs</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors clustered at the village level in parentheses. All columns are loan level OLS regressions. The outcome variable is whether the money lender demands a restrictive form of collateral, using the data driven definition of restrictive. Controls include number of agricultural cooperatives, whether the village has common land, and the education of the village head as well as their interaction with post. Income and loan size are trimmed at the 99th percentile. Only villages with between 50 and 250 households are included.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Figure 2-1: Patterns of Default by Wealth Level

Figure 2-2: Probability Household Borrows from Money Lender by Wealth Level
Figure 2-3: Probability Money Lender Uses Restrictive Collateral by Wealth Level

Figure 2-4: Coefficients on \( invsize_t \ast year_t \)
Bibliography


3.1 Introduction

The innovation of microfinance has long been regarded as the singular promising approach to large-scale poverty reduction, based on the belief that expanding credit among the poor would promote entrepreneurial activity and in turn lead to growth. However, enthusiasm for micro lending has begun to erode. While it has spread to over 100 countries and now serves more than 130 million borrowers (Microcredit Summit Campaign), there has been little systematic evidence that microfinance has a significant impact on poverty alleviation. Randomized control trials of micro-credit expansions find little impact on borrower income and business size (Banerjee, Karlan, and Zinman, 2015). Perhaps more surprisingly, cash grant experiments show that, despite the existence of microfinance institutions, micro-entrepreneurs still effectively face severe credit constraints as demonstrated by very high marginal returns to capital. (de Mel, McKenzie, and Woodruff, 2008).

Why has microfinance failed to live up to its promise? In this paper, we shed light on this question by developing a model that captures two empirical regularities of the industry’s rapid expansion—an increasing number of lenders available in a given market, and that entry of these overlapping lenders largely preceded the development and strengthening of the institutions that are thought to facilitate effective financial contracting. While such lender saturation or competition is normally associated with well functioning credit markets, we find that in the presence of weak credit market institutions, expanding financial access by introducing multiple sources of credit may severely backfire. Capital

* This chapter is joint work with Daniel Green.
allocation can be distorted away from the most productive use, which in turn generates perverse incentive for micro-entrepreneurs to choosing the least ambitious production technology.

Empirically, lender saturation and weak contracting environments are are increasingly associated with the narrative of a multiple borrowing crisis (e.g. see Faruque and Khalily, 2011). The quintessential example of multiple borrowing is the spectacular boom and bust of the microfinance industry the Indian state of Andhra Pradesh. A nascent industry in the 1990s, allegedly thousands of lenders entered the market to supply credit to nearly the entire state, much of it in the form of “overlapping” loans from many lenders to the same borrower.\(^1\) Borrowers accumulated large debt balances from many lenders, who individually had no way of observing or controlling a borrower’s total indebtedness. This ultimately proved unsustainable and culminated in a default crisis and near-collapse of the industry in 2010. Fear and realization of such multiple borrowing crisis have also arisen in Peru, Guatemala, Bolivia (de Janvry, Sadoulet, McIntosh, Wydick, Luoto, Gordillo, and Schuetz 2003), Uganda (McIntosh, de Janvry, and Sadoulet 2005), and Bangladesh (Faruque and Khalily 2011) following the rapid entry of new microfinance lenders.

In each of these cases, lenders, policymakers, and academics have all noted the importance of a specific implication of institutional weakness—the inability of borrowers to commit to an exclusive relationship with a single lender. Such commitment requires effective institutions that allow lenders can verify exclusivity (credit registries) and enforce punishments for any deviations (courts). If borrowers are unable to commit to a single lender ex-ante, then ex-post commitment induced by low marginal returns to further borrowing will be valuable in obtaining agreeable financing terms and facilitating investment. This logic suggests a nuanced relationship between financial development and resource allocation. Our model formalizes this insight. When borrowers cannot commit to exclusive contracts, better projects can receive less investment, borrowers may choose to forgo the most productive investments in favor of economically inferior alternatives, and that the severity of these distortions is increasing in the number of lenders available to borrowers. Together, these results outline a new explanation of why increased access to finance does not always improve aggregate outcomes.

In the model, an entrepreneur visits multiple lenders to obtain funds for a new investment opportunity. Crucially, the entrepreneur lacks the ability to commit to exclusive borrowing from a single lender and cannot write loan contracts that are contingent on the terms of contracts subsequently signed with other lenders. The entrepreneur faces an uncertain cost of default that is realized when loans come due, and defaults if the debt owed exceeds this cost. Thus, the more debt owed the less likely it is to be repaid. This gives rise to an externality between lenders. New lenders willingly provide additional investment that existing lenders would not, because new lenders do not internalize the decreased likelihood of repayment of existing debt when pricing new debt contracts.

\(^1\)At the peak in 2009, surveys estimated that 84% of rural villagers in Andhra Pradesh had loans from multiple lenders, including 58% borrowing from at least four lenders. See Johnson and Meka (2010), Taylor (2011) and references therein for a collection of facts and anecdotes about multiple borrowing in Andhra Pradesh.
Rational lenders anticipate this additional borrowing and offer loan terms that compensate them for it, making multiple borrowing undesirable ex-ante, but without commitment unavoidable ex-post. Inability to commit to an exclusive lending relationship is thus a binding constraint and in equilibrium induces distortions in lending and investment outcomes.

The economics of the commitment externality highlighted above were first demonstrated in the seminal work of Arnott and Stiglitz (1991, 1993) and applied to credit market settings by Bizer and DeMarzo (1992). To help build intuition for how the commitment externality operates in our setting, we show that their main findings, that lack of commitment increases interest rates and debt accumulation, also hold in the investment context of our dynamic model. These implications of lack of commitment echo the observed high levels of debt and high interest rates in microfinance markets where borrowers can obtain loans from many lenders simultaneously but contracting sophistication is limited.

We then show that distortions to investment induced by commitment problems depend crucially on the nature of the new investment opportunity the entrepreneur is attempting to finance, and find that more is not always better. More efficient opportunities may receive less investment and face higher interest rates, and the most promising investment opportunities may not be undertaken at all. These results together introduce a novel microfoundation for the relationship between credit market imperfections and aggregate resource misallocation.

At the heart of these results is that lower marginal returns create endogenous commitment power not to borrow (much) from additional lenders—the benefits of marginal investment return are not as attractive relative to the cost of higher debt obligations. Since entrepreneurs cannot commit ex-ante to limit inefficient future borrowing, properties of their future marginal returns to investment that induce them to limit such borrowing ex-post are especially valuable. This induces a trade-off between an investment technology’s efficiency and concavity. Low marginal (and hence average) returns are bad for output, but declining marginal returns are good for incentives. When marginal returns are high, borrowers who lack commitment not to continue borrowing may end up on the wrong side of the “debt Laffer curve,” meaning they are receiving an amount of investment that could have been supported by a lower quantity of debt, if only they could have committed to it. Financial constraints are thus endogenously more severe for better projects when they also have sufficiently higher marginal returns.

This relationship between financial constraints and productivity generates a particularly stark form of misallocation: a negative correlation between the level of investment in a project and its productivity. In the presence of commitment problems, projects that warrant the most investment end up receiving the least. This is consistent with a growing body of evidence from micro and experimental studies conducted in developing economies that productive firms may be especially credit constrained. McKenzie and Woodruff (2008) study microenterprise in Mexico and use a randomized experiment to estimate returns to investment capital. In addition to finding very high
returns on average, they also find a positive relationship between the returns to investment capital of borrowers and the financial constraints they face, indicating that it is the projects with the best economic fundamentals that are most under-served. Similarly, Banerjee and Duflo (2014), using policy changes in credit access programs in India, find similar evidence of a negative selection effect. In their paper, the ordinary least squares relationship between loan growth and profit growth has a lower magnitude than the causally estimated specification, which they conclude implies that it is the least productive firms that are acquiring the most credit. Our paper shows that such misallocation can be explained by commitment problems in credit markets.

Next, we endogenize project choice by the entrepreneur to reveal the full extent of the misallocative forces of multiple borrowing and lack of commitment. Not only can lack of commitment cause better projects to receive less investment, but it can also cause entrepreneurs to reject the most profitable projects entirely in favor of less efficient endeavors. When commitment problems are severe, constrained entrepreneurs choose business plans that have low prospects for expansion precisely because these are the easiest to finance. Those who instead choose better projects will receive low levels of funding at high interest rates, reducing their ability to accumulate equity to finance expansion.

Turning to policy, we show that optimal policy in the face of commitment problems strongly mirrors India's regulatory response to the Andhra Pradesh crisis. In 2011, the Reserve Bank of India imposed new regulations on the banking industry and stated that they were in part meant to address multiple borrowing. A debt limit (USD 790) and interest rate cap (26%) were imposed. The Malegam Report (2011) also recommended borrowers could only have loans from at most two microfinance lenders simultaneously, but this was not implemented in the final regulations. Each of these policies works by explicitly limiting the impact of lack of commitment in lending markets. However, these policies are all imperfect. If borrowers could commit, or incentivize themselves to commit through contingent contracting, the externalities that arise between lenders would disappear and outcomes would improve, including the efficiency of investment allocation across projects.

The emergence of exactly this sort of contingent contracting in sophisticated financial markets thus further substantiates the idea that commitment distortions are important. In the syndicated loan market in the United States and Europe, widely employed “performance pricing” covenants allow interest rates to vary based on changes in observable firm characteristics that occur after the loan has been issued. Making interest rates increasing in a firm’s total amount of debt, as the majority of these covenants dictate, internalizes the spillovers between lenders and restores the full-commitment lending outcomes, even when borrowers cannot explicitly agree to exclusive borrowing. Of course, implementing such contracts requires lenders can observe violations (for example through credit registries) and contract on them (for example through courts).

\footnote{It was also stated that at least 75% of the total loans originated by any regulated MFI should be for the purpose of income generation. (Reserve Bank of India (2011))}

\footnote{See Asquith, Beatty, and Weber (2005) for a detailed description of performance based pricing covenants.}
Theoretically, our model is based on the theme of common agency, which describes environments in which multiple principals with possibly conflicting interests act to influence the behavior of a single agent. Bernheim and Whinston (1986) formalize the notion of common agency and obtain results on the efficiency of the incentive structures required to implement the equilibrium action and the optimality of the action that prevails in equilibrium. Segal (1999) generalizes many common agency models in a unifying framework and shows that when contracting is not publicly observable inefficiencies occur when there are externalities at the efficient outcomes because the agent cannot compensate the principals for the externalities they impose on each other. In our setup, the principals can be thought of as the multiple lenders, each with a limited ability to control total borrowing of the borrower, the agent.

Early applied work on common agency is a collection of papers by Arnott and Stiglitz (1991, 1993), which study common agency in insurance markets. These papers recognize that additional insurance providers can impose externalities on each other through moral hazard of the buyer. Bizer and DeMarzo (1992) bring this idea to credit markets. They study a model in which a borrower with a desire to smooth consumption between two periods can sequentially visit lenders to obtain loans backed by the borrower’s stochastic next-period income, which is influenced by non-contractible effort choice. As in the Arnott-Stiglitz series of papers, moral hazard generates an externality that new lenders offer loans that harm previous lenders by decreasing their expected profits. More recently, the “maturity rat race” of Brunnermeier and Oehmke (2013) highlights that commitment externalities can also arise in the context of maturity structure choice. There, it is the inability to commit to maturity structure, not the level of debt, that generates problems. Our main contribution to this literature is that we are the first to link commitment problems in credit markets to investment decisions. We show that commitment externalities cause better projects to receive less investment and encourage entrepreneurs to choose projects with low growth potential and rapidly diminishing returns to scale. Thus, our model provides a rationale for the failure of microfinance to make a significant impact on poverty and suggests that microfinance can only be successful if it is backed by the high quality institutions that support intermediation in the developed world.

It is also important to emphasize that our paper relates to competition in credit markets only through competition’s exacerbation of commitment externalities. We focus on the fact that easier access to additional lenders makes it harder to enforce exclusive borrowing and thus ex-post commitment mechanisms, such as selecting projects with limited returns to scale, are valuable. This channel is independent of the relationship between competition and market power in the context of imperfect credit markets, which has been studied in other papers. Of this work, the most related to our paper is Parlour and Rajan (2001), which shows that commitment problems prevent lender entry from competing away all lender profits. Also related is the theory of relationship lending.

---

developed by Petersen and Rajan (1994, 1995). This theory suggests that when borrowers cannot commit to long relationships with a single lender, limited market power of lenders erodes their ability to subsidize lending to early stage firms and recoup expenses later through rent extraction.

The rest of the paper is organized as follows. Section 2 describes our model and section 3 defines and solves for the equilibrium. In section 4 we show the implications of lack of commitment in credit markets with multiple lenders on interest rates and default rates, and derive our main result that entrepreneurs with more attractive investment opportunities can endogenously face more severe credit constraints and thus invest less than those with inferior projects. We illustrate the implication on inefficient project choice in section 5. We next examine, in section 6, policies adopted by regulatory agencies to address the multiple borrowing problem through the lens of our model. Section 7 concludes. Proofs of all propositions and lemmas are relegated to the Appendix.

3.2 Model

Overview and Timing

An entrepreneur attempts to finance a new investment opportunity with the constraints that it cannot commit to exclusive borrowing from a single lender and that there is limited enforcement of debt repayment, i.e., the borrower only repays if the costs of default are high enough. The model has two stages \((s = 1, 2)\) and contains two sets of agents: a single entrepreneur and an infinite sequence of potential lenders. All parties are risk neutral and have no discounting. The entrepreneur is endowed with a variable-scale investment opportunity that returns \(R(I)\) deterministically at \(s = 2\) for any \(I > 0\) invested at \(s = 1\). While the returns to this investment opportunity are observable, they are not perfectly pledgeable. After the project's returns are realized, the entrepreneur learns of its cost of defaulting on its debt, and only repays if the default cost exceeds the amount of debt owed. Specifically, we assume default costs \(\tilde{c}\) are drawn from a distribution \(F_{\tilde{c}}\). Debt \(D\) is repaid if \(\tilde{c} > D\), and we denote the probability of repayment as \(p(D) = 1 - F_{\tilde{c}}(D)\), which is of course a decreasing function of \(D\). Figure 3-1 summarizes the timing of the model.

We model the entrepreneur raising capital from the lending market at \(s = 1\) as an infinite horizon dynamic game of complete information in which the entrepreneur sequentially visits a potentially infinite number of lenders. All lenders are risk neutral and do not discount between \(s = 1\) and \(s = 2\). Their opportunity cost of funds is the risk-free rate, which we normalize to one throughout the paper. Upon meeting a lender, the entrepreneur makes a take-it-or-leave-it offer for a simple debt contract \((L_i, D_i)\) specifying the amount borrowed \(L_i\) and the promised repayment amount \(D_i\). The lender, who can observe the history of the entrepreneur's borrowing from previous lenders, chooses whether to accept or reject this offer, and if accepted the funds are exchanged.

At this point, the entrepreneur loses access to the lending market with probability \(1 - q\), with
Entrepreneur visits the lending market and leaves borrowing $L$ and promising to repay $D$.
Invests $I = L$ into the new investment opportunity.

Output $R(I)$ realizes.
Default cost $\tilde{c} \sim U[0, 1]$ is drawn.
Entrepreneur repays debt if $D \leq \tilde{c}$.

Figure 3-1: Timing of the Model

- Lending game starts
- Entrepreneur makes proposal
- Lender 1 accepts / rejects
With probability $q$:
- Entrepreneur meets lender 2
- Lender 2 accepts / rejects
With probability $(1 - q)$:
- Entrepreneur invests total amount raised
- Game moves on to stage 2
With probability $q$:
- Process repeats
- Entrepreneur meets exactly K lenders with probability $(1 - q) \cdot q^{(K-1)}$
- Entrepreneur invests total amount raised
- Game moves on to stage 2

Figure 3-2: The Lending Market Game at $s = 1$

$q \in [0, 1)$. Otherwise, the entrepreneur meets a new lender and the process described above is repeated until eventually access to the lending market is lost or the last lender is visited. By assumption, no lender is visited more than once. After losing access to the lending market, the entrepreneur invests the aggregate financing it raised in the new project, and the model progresses to stage $s = 2$. Figure 3-2 provides a graphical exposition of the lending market game of stage $s = 1$.

A brief discussion of the important aspects of the model is in order. The crucial feature of the model is that the entrepreneur cannot credibly commit ex-ante to avoid borrowing from subsequent lenders it meets. More broadly, debt contracts cannot be contingent on the terms of other debt contracts written with subsequent lenders. This form of contractual incompleteness precisely reflects, in our view, important features of the institutional and contracting environments in which we think misallocation and suboptimal technology choice are most salient.

The assumption of sequential borrowing from a possibly large number of banks need not be taken literally. What drives our results is that lack of commitment to an exclusive lending relationship
leaves room for the externalities between lenders and perversely affects equilibrium outcomes. Our results also hold when the entrepreneur borrows simultaneously, rather than sequentially, from a finite number of lenders where loan terms cannot be made contingent on the other loans taken by the borrower. However, such a model is analytically less tractable and subject to the standard critique in the simultaneous contracting literature that the results are sensitive to the specification of agents' off-equilibrium beliefs (Segal and Whinston (2003)).

We formulate the endogenous choice of default via a random default cost that is realized ex-post. The formulation is similar to a random shock to an outside option, as in Aguiar, Amador, Hopenhayn, and Werning (2016), and reflects the fact that forces outside of the model generate ex-ante indeterminacy in the ex-post costs of defaulting on debt. Weak institutions are a salient driver of such indeterminacy; in the Andra Pradesh default crisis, repayment rates plummeted from near-perfect to near-zero almost overnight when grandstanding local politicians urged borrowers to stop repaying their debts. In the appendix, we provide a microfoundation of the random default through moral hazard.

The parameter \( q \) exogenously limits the amount of commitment power borrowers can obtain in the lending market. The closer \( q \) is to zero, the less likely it is that contingencies arise in which early lenders can be exploited by further borrowing. There are several appealing interpretations of this parameter. First, \( q \) can be thought of as inversely related to the difficulty or cost of subverting commitment, and reflects the quality of the contracting environment. As \( q \to 0 \) the contracting environment is able to perfectly enforce contingency in loan terms. When \( q \to 1 \) borrowers can costlessly find new lenders to provide marginal lending. A second view is that \( q \) reflects the composition of search frictions in the lending market and the limited time an entrepreneur has to raise money for an investment opportunity. Under this interpretation one would assume a complete inability to conduct contingent contracting or exclusive borrowing--the severity of the commitment problem is determined by the market structure of lending and time preference for funding.

### 3.3 Equilibrium

To begin solving the model, we introduce two simplifying assumptions which aide greatly in the exposition of the model and in highlighting the underlying economic forces generating our results. 

**Assumption 3.1.** *The return on the new investment opportunity is linear and productive:*

\[
R(I) = \alpha I \quad \text{and} \quad \alpha > 1.
\]

**Assumption 3.2.** *Default costs for the entrepreneur are uniformly distributed between zero and one.*

\[
\bar{\epsilon} \sim U[0, 1].
\]
Linearity in returns to the investment opportunity means that the level of investment financing obtained from prior lenders is not directly relevant for the objective of the borrowers or lenders at any stage of the lending game, greatly simplifying the model. We restrict to efficient linear projects (\( \alpha > 1 \)) to rule out trivial cases in which there is no demand for borrowing or investment. The uniform distribution of default costs between 0 and 1 normalizes the borrower’s maximum debt capacity to one and generates simple expressions for properties of debt repayment. The probability of repayment is now simply \( p(D) = 1 - D \) and the expected debt servicing cost (either repayment or strategic default) is \( \mathbb{E} \left[ \min(\bar{e}, D) \right] = D - D^2 / 2 \).

**The Full Commitment Case**

We now characterize the solution to the model when borrower can commit to visiting a single lender and use this as a benchmark to illustrate how lack of commitment affects outcomes in the model. The ability of the entrepreneur to commit to borrowing from a single lender can be captured by setting \( q = 0 \), since this ensures it is common knowledge that the borrower will visit exactly one lender.

The above model now simplifies substantially. The entrepreneur makes a take-it-or-leave-it offer \((D, L)\) to the lender and invests \( I = L \) in the new investment opportunity. The lender accepts any loan offer that is weakly profitable in expectation. The optimal lending contract is found by maximizing the entrepreneur’s welfare subject to the lender participation constraint. Denote this problem \( P^{SL} \) for the “single lender” problem:

\[
(P^{SL}) \quad \max_{D, I} R(I) - \mathbb{E} \left[ \min(D, \bar{e}) \right]
\]

\[
\text{s.t. } I \leq p(D) D \quad \text{(lender's participation constraint)}
\]

**Proposition 3.1.** Under Assumptions 1 and 2, the solution to the single lender problem \( P^{SL} \) is characterized by the condition

\[
\alpha \times \left[ p \left( D^{SL} \right) + p' \left( D^{SL} \right) D^{SL} \right] = p \left( D^{SL} \right)
\]

At the optimum, the costs and benefits of pledging an additional dollar of face value of debt must be equalized. The marginal cost is the increase in expected debt repayment costs associated with the additional borrowing, which is simply the probability of actually repaying the marginal dollar pledged.\(^5\) The gain from pledging an additional dollar of face value of debt is the marginal return to investment times the marginal investment that can be raised from pledging an additional dollar of face value of debt. Because extra repayment reduces the value of all debt claims (through

\(^5\)One must also consider the decrease in expected repayment probability and the increase in expected default costs conditional on default. However, these terms exactly offset each other—at the margin the borrower is simply replacing some physical repayment with increased default costs.
limited enforcement), the value of existing debt falls when new debt is issued. Because in this example there is only a single lender, the lender internalizes the change in value of all existing debt coming from the marginal issuance of new debt, and thus can make zero profits by supplying \( p(D) + p'(D)D \) of investment for a marginal dollar of promised repayment. We now highlight the externality that arises when the entrepreneur can obtain loans from more than one lender.

Illustrating the Commitment Problem

When \( q > 0 \) the borrower cannot obtain the same loan contracts from the first lender that it could when \( q = 0 \). The first lender now anticipates the possibility that the borrower will obtain additional debt from future lenders that dilutes the value of its original debt claims and thus cannot profitably accept the same loan terms as in the full commitment case. To illustrate why this problem arises, imagine the borrower can borrow sequentially from two lenders, but cannot commit to deal exclusively with one lender or to make terms of either loan conditional on those of the other. Therefore, loan terms with the second lender are arranged to maximize the surplus to the borrower and the second lender taking the terms of the first loan as fixed.

In this example the second lender imposes an externality on the first whenever the first lender has issued a positive amount of debt. To see this, denote the face value of debt promised to the first lender by \( d_1 \) and the face value of debt promised to the second lender by \( d_2 \) and define \( D = d_1 + d_2 \) to be the total face value of debt. Subject to the same regularity conditions as above, taking aggregate debt \( D \) as given the following first order equation characterizes \( d_2 \), the amount of debt issued to the second lender.

\[
\alpha \cdot [p(D) + p'(D) \cdot d_2] = p(D).
\]

This is identical to Equation 3.3.1 except that the term in square brackets is different. Here, the second lender only internalizes the change in the value of its own debt due to its marginal debt issuance at the optimum, not the change in the total value of debt issued. Thus the second lender imposes an externality on the first. When borrowers lack commitment lenders must anticipate future borrowing and charge higher interest rates that compensate them for value they expect to lose. This is exactly why the commitment problem changes lending outcomes in our model and is the key intuition behind the results in Bizer and DeMarzo (1992) that multiple lenders lead to higher interest rates and higher borrowing than would be obtained if commitment were possible. Now that we have clearly illustrated why lack of commitment affects investment outcomes and technology choice.
Equilibrium without Commitment

We now turn to studying the full model without commitment. Our solution concept is Markov Perfect Equilibrium (MPE) in which the borrower's strategy is a function only of a single state variable $D$, the cumulative face value of debt the issued so far in the game, and the lender's strategy is a function of $D$ as well as the current loan proposed by the borrower. Modeling borrowing as sequential, rather than simultaneous, allows us to avoid issues of the sensitivity of our results to off-equilibrium beliefs. As we show at the end of this section, the equilibrium we identify captures the limiting outcome of the unique subgame perfect equilibria of finite-lender versions of the model as the number of lenders grows large. We now provide a formal definition of a Markov Perfect Equilibrium in the lending market game.

**Definition 3.1.** A Markov Perfect Equilibrium (MPE) of the infinite lender lending market game is a set of borrower and lender strategies that are mutual best responses at every subgame when subject to the following constraints.

a) The entrepreneur's strategy when encountering any lender is a mapping from how much it has already pledged to repay, $D$, to a simple debt contract $(L_i, D_i)$.

b) All lenders' strategies are represented by functions mapping from the state variable $D$ and the loan contract proposed by the entrepreneur to the lender's decision to either accept or reject the proposal.

We begin characterizing the equilibrium of this game by exploring the best responses of each player.

**Lender Best Responses**

Lenders have rational expectations of future borrowing and thus only accept contracts that yield non-negative expected returns taking the strategies of the borrower and other lenders as given. A lender receiving a loan proposal must, given the state of the game, evaluate the expected profit from the loan taking into account the strategies of the borrower and future lenders and the likelihood that the borrower will be able to meet these lenders at all. Specifically, define $D' \equiv D + D_i$, and denote $\bar{p}_i(D')$ to be lender $i$'s perceived probability that the entrepreneur will not default on a loan $(L_i, D_i)$ to this lender, given the borrower has previously obtained total face value of debt $D$. Such a loan is weakly profitable in expectation if $L_i/D_i \leq \bar{p}_i(D')$. Since lenders are risk neutral and do not discount stage 2 cash flows, lender $i$'s optimal strategy is to accept the loan offer if and only if it is weakly profitable. We focus on a symmetric equilibrium in which lenders' best response can be characterized by a function $\bar{p}(\cdot)$ such that $\bar{p}_i(\cdot) = \bar{p}(\cdot)$ for all lenders $i$. 

213
Borrower Best Response

Conditional on meeting a lender, and given lender strategies represented by $\bar{p}(\cdot)$, the borrower makes a loan offer that maximizes its expected continuation utility, taking into account the chances of being able to meet more lenders and obtaining further marginal borrowing. Solving for the entrepreneur's best response function thus involves a dynamic optimization problem. Taking lender pricing as given the entrepreneur forms strategies that maximize it's continuation utility at each value of the state variable $D$.

The borrower knows that lenders will accept any loan they expect to be profitable, so to maximize its own utility it will only offer loans that lenders expect to make exactly zero profits. Thus given the total face value of debt borrowed so far $D$, the borrower only proposes loans $(L_i, D_i)$ that satisfy $L_i = \bar{p}(D') D_i$, where again $D' \equiv D + D_i$. Upon meeting any lender, if the entrepreneur has cumulative debt $D$ and leaves the lender with cumulative debt $D'$ then the maximum amount of new investment the lender would provide is given by $\bar{p}(D') [D' - D]$. The lender expects the loan to be repaid with probability $1 - q$, in which case it will be return $D' - D$. Thus, taking loan pricing as given the entrepreneur solves:

$$V(D) = \max_{D'} \alpha \bar{p}(D') (D' - D) - (1 - q) \mathbb{E} [\min (D', \tilde{c})] + qV(D') \quad (3.3.3)$$

Conditional on arriving at a new lender having already issued face value of debt $D$ the entrepreneur optimally chooses $D'$, the new total face value of debt it will have issued after contracting with this lender. The first term on the right hand side of Equation 3.3.3 is the marginal payoff from the new investment opportunity associated with obtaining additional investment $\bar{p}(D') (D' - D)$. With probability $1 - q$ the entrepreneur loses access to the lending market and either repays debt $D'$ or defaults and pays the default cost $\tilde{c}$ if it is lower than the cost of repayment. Finally, with probability $q$ the entrepreneur does not lose access to the lending market and will receive continuation utility $V(D')$ from future borrowing. Plugging in for $\mathbb{E} [\min (D', \tilde{c})]$ given default costs $\tilde{c}$ are uniform from zero to one gives:

$$V(D) = \max_{D'} \alpha \bar{p}(D') (D' - D) - (1 - q) \left( D' - \frac{D'^2}{2} \right) + qV(D') \quad (3.3.4)$$

Denote a policy function that solves the dynamic programming problem in Equation 3.3.4 by $g(D)$. Conditional on arriving to a new lender with aggregate face value of debt $D$, the borrower will leave with aggregate face value of debt $D' = g(D)$, having proposed a new additional loan $(\Delta D, \Delta L)$ with $\Delta D \equiv g(D) - D$ and $\Delta L \equiv \bar{p}(g(D)) (g(D) - D)$. On path, following this strategy generates a sequence of total aggregate face values of debt that have been accumulated up to a given lender, conditional on the lending market progressing that far: $\{g(0), g(g(0)), g^3(0), \ldots\}$. The aggregate face value of debt obtained in the lending market is thus a random variable $D_{agg}$ that realizes a
particular value of this sequence depending on how many lenders the borrower is able to visit before
the lending game ends.

**Equilibrium Characterization**

Given the discussion of borrower and lender best responses above, it is clear that a MPE of
the model is equivalent to a solution a fixed point dynamic programming problem. Given lender
strategies (captured by \( p(-) \)), the borrower’s optimal strategy is represented by a policy function
\( g(-) \) that solves the dynamic programming problem in Equation 3.3.4. Given the distribution of
total aggregate face value of debt \( D^{agg} \) induced by \( g(-) \), each lender forms rational expectations over
the probability the borrower will repay, denoted by \( \tilde{p}(D') = E[p(D^{agg})|D'] \). A set of strategies
forms a MPE if lender strategies given by \( \tilde{p}(-) \) are rational given \( g(-) \), and these borrower strategies
are optimal given lender strategies embodied in \( \tilde{p}(-) \). We now summarize this characterization of
an MPE in the following Theorem.

**Proposition 3.2.** A symmetric Markov Perfect Equilibrium of the lending game is characterized
by functions \( \tilde{p}(-) \) and \( g(-) \) that map from cumulative debt level \( D \in [0, 1] \) to the interval \( [0, 1] \) such
that:

a) \( g(-) \) is the policy function in the solution to the dynamic programming problem in Equation
3.3.4 taking \( \tilde{p}(-) \) as given.

b) Lenders’ perceived expected repayment probabilities \( \tilde{p}(-) \) used to form accept/reject strategies
are correct taking \( g(-) \) as given.

\[
\tilde{p}(D) = E[p(D^{agg})|D] = 1 - E[D^{agg}|D]
\]

**Closed Form Solution**

We now solve for the unique linear symmetric MPE in closed form. Notice that if \( \tilde{p}(D) \) were
linear then the dynamic programming problem would have a linear-quadratic form, so we look for
an equilibrium where \( \tilde{p}(D) \) and \( g(D) \) are linear functions of \( D \). The following lemma characterizes
the form of the solution to this problem.

**Lemma 3.1.** For \( \alpha > 1 \) and \( 1 > q > 0 \), there exists \( \ell^* (\alpha, q) \geq 1 \) and \( b^* (\alpha, q) \geq 1 \) such that the
unique linear MPE takes the following form:

\[
\begin{align*}
\tilde{p}(D) &= (1 - D) \cdot \frac{1}{\ell^*} \\
1 - g(D) &= (1 - D) \cdot \frac{1}{b^*}
\end{align*}
\]

where \( \ell^* \) and \( b^* \) respectively parameterize lender and borrower strategy in equilibrium.
There is an intuitive interpretation for both $\ell^*$ and $b^*$. First, if the borrower arrives to a lender with current debt $D$, its remaining debt capacity is $1 - D$. When leaving this lender the borrower will have pledged a total of $g(D)$ and thus the borrower will have remaining debt capacity $1 - g(D)$. Therefore $\frac{1}{b^*}$ is the fraction of current borrowing capacity that remains after visiting a lender. A higher $b^*$ (or lower $\frac{1}{b^*}$) corresponds to more aggressive borrowing by the borrower; it will deplete its available debt capacity more rapidly. Given $g(D) > D$, in equilibrium the entrepreneur borrows a positive amount from each lender it gets to visit, no matter how much it has already borrowed.

Second, recall that with commitment, the repayment probability is exactly $\bar{p}(D) = p(D) = 1 - D$, which corresponds to $\ell^* = 1$. Thus $\ell^* > 1$ corresponds to lower expected repayment probability and higher interest rates. In other words, when $\ell^*$ is high the lenders make pricing decisions as if they expect the borrower to accumulate substantially more debt from future lenders, which deteriorates the value of the current lender’s own claims, and set interest rates to reflect this.

Finally, there is also a static interpretation for the dynamic equilibrium. Taking lender’s strategy parameterized by $\ell$ as given (which determines $\bar{p}(D)$ and the interest rates), the entrepreneur chooses how much debt to issue when meeting each lender. This specifies a best response $b = B(\ell)$ for the entrepreneur, which can be thought of as representing a loan demand schedule. Since higher interest rates induce the borrower to take out debt less aggressively, the loan demand schedule is downward-sloping ($B'(\ell) < 0$). On the other hand, given the aggressiveness of entrepreneur’s borrowing behavior parameterized by $b$, the lenders are able to determine the distribution of the face value of total borrowing. This maps into the distribution of the value of their own debt claims, and lenders set their decision rule such that they only accept loans they expect to be weakly profitable, and thus specifies a best response $\ell = L(b)$ as the loan supply schedule. The interest rates lenders have to charge in order to break-even increases as the entrepreneur takes out loans more aggressively, hence the loan supply schedule is upward-sloping ($L'(b) > 0$). The unique equilibrium is then the unique intersection of the “loan demand” equation $B(\ell)$ and the “loan supply” equation $L(b)$, which is depicted in Figure 3-3.

**Equilibrium Choice**

We choose to focus on the unique linear symmetric MPE in our infinite lender model because it is closely related to the unique SPE of finite lender version of our game. Because the number of lenders in any real world credit market is finite, our infinite lender model and the linear MPE solution concept is merely an abstraction that comes with the benefits of algebraic tractability.

---

6 If the borrower acquired additional debt of more than $1 - D$ it would default on all debt for sure. Recall default costs $\bar{c} \sim U[0, 1]$, so if the aggregate face value of debt equals or exceeds one the debt will never be repaid, as it will be less costly to default on it no matter what realization of default costs occur.
Formally, we define the finite \(N\)-lender game by modifying our infinite lending game as follows. After meeting the \(i\)-th lender, the borrower gets to meet \((i+1)\)-th lender with probability \(q\) if and only if \(i+1 \leq N\), and with probability zero otherwise. That is, we truncate the game at a maximum of \(N\) lenders, while keeping the stochastic nature of the lending game unchanged for the first \(N\) lenders. Note that the finite \(N\)-lender game admits a unique \(SPE\) which can be solved by backward induction. The following proposition demonstrates that the unique linear symmetric \(MPE\) in the infinite lender game can be obtained as the limit of the sequence of the unique \(SPEs\) of the finite lender games as we take the number of lenders to infinity.

**Proposition 3.3.** Fix \(i\). In the finite \(N\)-lender game with \(N > i\), as \(N \to \infty\) the strategies at lender \(i\) in the unique \(SPE\) converge uniformly to the corresponding linear symmetric \(MPE\) strategy of the infinite-lender game.

### 3.4 Comparative Statics

We now proceed by highlighting the role of commitment in equilibrium investment outcomes. First, because the outcome of the lending game depends on the random number of lenders the borrower meets, we will focus our attention on expected outcomes, such as the expected face value of debt, the expected level of investment, the expected aggregate interest rate, and expected welfare. Due to the linear strategies in the game these concepts have simple closed form expressions. We first explore comparative statics of these outcomes to the model parameters \(q\) and \(\alpha\), and then turn to broader implications outside the model of how these forces can endogenously affect project choice by tightening financial constraints.
3.4.1 Comparative Statics of Commitment

How does the exogenous severity of the commitment problem \( q \) impact equilibrium outcomes? The expected number of lenders a borrower visits is \( \frac{1}{1-q} \) which increases with \( q \). Therefore a higher \( q \) can be seen as a worsening of the commitment problem. The following result echoes what has been experienced in the microfinance industry by a broad range of developing markets: the availability of lenders is associated not only with multiple borrowing, but also higher levels of debt, higher interest rates, and lower repayment rates.

**Proposition 3.4.** The following results describe the equilibrium effects of increasing access to additional lenders, as parameterized by \( q \), for any \( \alpha > 1 \) and \( q > 0 \):

- **a)** Expected aggregate face value of debt \( (E[D^{agg}]) \) increases in \( q \).
- **b)** Expected investment \( (E[I^{agg}]) \) decreases in \( q \).
- **c)** Expected probability of default \( (E[p(D^{agg})]) \) increases in \( q \).
- **d)** The ex-ante expected interest rate \( (E[D^{agg}] / E[I^{agg}]) \) increases in \( q \).
- **e)** Ex-ante welfare of the entrepreneur decreases in \( q \), and in the limit as \( q \to 1 \) welfare converges to zero, the level that would be obtained if the entrepreneur does not have access to the lending market at all.

As \( q \) increases the expected face value of debt raised in equilibrium increases. This is the net effect of two forces. First, for higher \( q \) the borrower's demand aggressiveness curve shifts down, meaning the borrower issues a smaller fraction of its debt capacity at each round of financing for a given interest rate strategy. This is because higher \( q \) means it is more likely that the borrower will be able to exploit the externality future lenders impose and obtain favorable pricing on subsequent loans. In effect, when \( q \) is higher the borrower wants to smooth borrowing over multiple lenders to obtain better aggregate financing terms. From the lender perspective, for a given borrowing strategy a higher \( q \) increases lenders' expectations of debt dilution, causing them to raise interest rates. This further reduces borrower demand aggressiveness. However, the net reduction in aggressiveness of debt accumulation is dominated by the increasing likelihood that more loans will be realized.

In other words, while lenders respond to concerns about debt dilution by raising interest rates, without commitment this cannot prevent increased expected borrowing. As summarized in Proposition 3.4, increasing \( q \) also increases the effective interest rate and probability of default, while expected investment falls. The possibility of multiple borrowing makes it harder to use available debt capacity to fund new investment.

Entrepreneur welfare (and hence total surplus) is also declining in \( q \). In fact, as \( q \to 1 \), the the commitment problem becomes so severe that while there is positive borrowing and investment, all surplus from investment is offset by costly default. The entrepreneur would be just as well off not
investing at all. This result is striking similar to the well-known Coase Conjecture in the industrial organization literature, that a durable good monopolist competing with itself intertemporally is unable to obtain any monopoly rents as consumers become very patient.\footnote{Specifically, the Coase Conjecture states that as consumers become extremely patient, a durable goods monopolist who cannot commit to future prices would have to sell the its good at competitive prices instantaneously and is unable to raise any profits. See Fudenberg and Tirole (Chapter 10, 1991).} Both results illustrate the fact that dynamic commitment problems can completely unravel the ability of an agent to capture or generate surplus. In our model, welfare strictly decreases with $q$ and at the limit the welfare level is as if the entrepreneur does not have access to a credit market at all. However, the welfare implications of lack of commitment for the monopolist (as in Coase Conjecture) are opposite—extreme patience drives monopolist rents to zero but maximizes social efficiency.

The predictions of Proposition 3.4 bear similarity to what is being observed in microfinance markets facing increases in competition among lenders. Most directly, McIntosh et al. (2005) find that increased competition between microfinance lenders in Uganda lead to increased debt accumulation and declining repayment rates of their borrowers. They cite conversations with local lenders that suggest that multiple borrowing is to blame:

"The chief executives of most of the [lending] institutions involved in this article were interviewed on the topic, and few were worried about competition insofar as it relates to the growth prospects of their institution. A common concern, however, was that, wherever two or more institutions are operating, many clients may be taking loans from several lenders simultaneously, or double-dipping...Nonetheless, they were unanimous in the opinion that this behavior does drive up default rates."

The most striking aspect of this anecdote is not that multiple borrowing occurs in the face of competition, but that lenders are very concerned about its effects on repayment and that its incidence cannot be controlled. Such anecdotes are abundant in the narrative surrounding multiple borrowing in microfinance.\footnote{A set of case studies of microcredit markets in Peru, Guatemala, and Bolivia (de Janvry et al. 2003) highlights the prevalence of multiple borrowing and high levels of debt that occurs following the rapid entry of lenders in each of these countries. Policy makers in Bangladesh are concerned that incidence of "overlapping borrowing" across multiple microfinance lenders is on the rise and incidence may be as high as 60% (Faruque and Khalily 2011). A survey of microfinance usage in Andrah Pradesh analyzed by Johnson and Meka (2010) suggests that borrowers in Andrah Pradesh borrow from multiple lenders within the same month because they cannot obtain enough financing from individual lenders.}

### 3.4.2 Comparative Statics of Project Returns

We now turn to the more surprising results about what happens when the marginal returns $\alpha$ of the investment opportunity increase. When returns to the investment opportunity are higher the borrower has a greater incentive to exploit the externality associated with the commitment problem and take on more new financing at the expense of previous lenders. A central result of our paper
is that for projects with higher returns, the effects of the commitment problem can be worsened to
the extent that the equilibrium may involve lower levels of investment than would occur with less
desirable projects.

To develop this result, first consider the effect of an increase in $\alpha$ on the equilibrium face value
of debt. As shown graphically in Figure 3-3, this increases borrowing aggressiveness—the borrower
shifts debt accumulation toward earlier lenders, increases the expected face value of debt. Why?
At any point in the game the borrower faces the following tradeoff. Taking interest rate strategies
as given, it weighs the benefits of borrowing more from the current lender for sure, or taking the
gamble that it will be able to meet a new lender from which to borrow marginally at better interest
rates that the previous lender would not have offered. When the returns to investment are higher
this tradeoff tilts away from a less aggressive borrowing strategy of waiting to try and exploit the
externality on the current lender.

Of course, more aggressive borrowing means lenders have to charge higher interest rates in
anticipation of higher aggregate borrowing. This result is summarized in Proposition 3.6.
Proposition 3.5. Expected face value of debt $E[D]$ and the effective interest rate $E[D] / E[I]$ are
increasing in $\alpha$, the returns to scale of the investment technology.

Now we turn to the main result on the relationship between the equilibrium level of investment
$E[I]$ and the returns to scale of the new project.
Proposition 3.6. Fix any $q \in (0, 1)$

a) The equilibrium level of investment $E[I]$ is non-monotone in $\alpha$. In particular, there exists a
cutoff $\alpha(q)$ such that expected investment is increasing in $\alpha$ below this cutoff, and decreasing
in $\alpha$ for $\alpha > \alpha(q)$. Formally,

$$\frac{dE[I]}{d\alpha} \begin{cases} < 0 & \text{for } 1 \leq \alpha < \alpha(q) \\ > 0 & \text{for } \alpha > \alpha(q) \end{cases}$$

b) The stronger the commitment problem in the lending market, the lower is the cutoff level of
marginal return $\alpha(q)$ for decreasing investment:

$$\frac{d\alpha(q)}{dq} < 0$$

These results can be visualized in Figure 3-4, which plots equilibrium expected investment as
a function of $\alpha$ for three different values of $q$. For all levels of $q$, the level of investment that the
entrepreneur gets to raise in expectation first increases in $\alpha$ and then decreases. Entrepreneurs with
better opportunities could be facing tighter constraints. The second part of the proposition shows
that such endogenous misallocation of resources is more severe in markets where the commitment
problem is worse.

220
The seemingly perverse outcome of lower investments in better opportunities arises because when commitment problems are present, it is possible that the equilibrium involves using available debt capacity inefficiently. To better understand the intuition behind this result, recall that the present value of promised repayments to creditors is affected by the face value of these claims in two directions. Holding repayment fixed, higher face value translates into higher present value. However, a higher face value of debt also reduces the probability of repayment. This gives rise to an aggregate debt Laffer curve: the value of debt is initially increasing in the amount of debt pledged starting from zero, but for sufficiently high face values of debt, pledging more debt actually results in a lower present value of debt because the negative force begins to dominate the positive one.

Figure 3-5 shows precisely how the debt Laffer curve arises in this model. This figure translates the best response curves of Figure 3-3 into the expected investment vs expected debt space by plotting the values of expected investment and debt that would result if the lenders were optimally responding to arbitrary strategies of the borrowers, and vice versa. This exercise generates two distinct curves, the (non-zero) intersection of which corresponds to the unique linear equilibrium allocation. The hump-shaped curve embeds lender optimality over a range of responses to all possible borrower strategies \( b > 1 \). This is the debt Laffer curve introduced in the previous paragraph. Given a borrowing aggressiveness strategy of the entrepreneur, there is an associated zero-profit level of expected investment.

---

9 An intersection of the curves in Figure 3-3 implies an intersection in Figure 3-5, but the converse is not true. This is because the mapping between \((b, \ell)\) and \((\mathbb{E}[D^{b,\ell}], \mathbb{E}[I^{b,\ell}])\) is not one-to-one: \((0,0)\) in the \((\mathbb{E}[D^{b,\ell}], \mathbb{E}[I^{b,\ell}])\) space is never an equilibrium of the model for \( \alpha > 1 \).
Lending Market Equilibrium: $q = 0.8$

![Graph showing the relationship between expected investment and expected debt across different values of $\alpha$.](image)

Figure 3-5: Expected Investment and Expected Debt

The upward sloping curves plot values of expected investment and debt if the borrower responds optimally to the range of all possible lender strategies $\ell \geq 1$. As illustrated in the graph, for higher $\alpha$ this curve shifts to the right, representing more aggressive borrowing by the entrepreneur upon meeting each lender. As marginal returns increase, the entrepreneur has more to gain by increasing borrowing (and investment). Importantly but unfortunately this is also a curse. Without commitment, the entrepreneur will also want to increase borrowing from any future lenders it gets to meet.

We know from Proposition 3.5 that $E[D]$ is monotonically increasing in $\alpha$. When marginal returns are high enough the equilibrium allocation will end up on the wrong side of the debt Laffer curve, where the entrepreneur is in equilibrium offering such high levels of repayment that it actually receives very little investment. Thus on the wrong side of the curve, increases in $\alpha$ increase expected debt but decrease expected investment. With the ability to commit to one lender, the borrower would never choose loan terms in this inefficient region.

These results are surprising and run against common intuition that more productive projects induce higher levels of investment despite the presence of financial constraints. In many classical theories of inefficient investment choice, investment occurs if and only if the net present value of a project exceeds a certain threshold, which can be above or below zero. Such models do not generate the stark patterns of resource misallocation that plagues developing countries. The commitment friction we explore in this paper however can generate these distortions. In fact, this force is so strong that it can also distort the influence the choice of investment opportunity chosen by entrepreneurs. In the next section we illustrate this point.
3.5 Endogenous Project Choice

Given the role of high marginal returns in Proposition 3.6, it is natural to ask what happens when project returns are not linear, but instead concave. In such a setting, borrowers can obtain commitment power by obtaining sufficiently large amounts of capital from early lenders such that the marginal returns to future borrowing, to which they cannot commit to avoid, are lower. Unfortunately, to study fully general production functions $R(\cdot)$ we must rely on numerical methods. However, we are able to analytically characterize an extreme form of concavity to illustrate our point. Consider investment opportunities have linear returns up to a certain size of investment, but then deliver zero marginal return for any additional investment beyond that point. We call these “linear-flat” projects and we use them to show that even if one investment opportunity is strictly dominated by another in terms of having lower returns for any level of investment, the dominated project could yield higher welfare for the entrepreneur. This result is summarized in Proposition 3.7.

**Proposition 3.7.** There exist pairs of projects $R_1(\cdot)$ and $R_2(\cdot)$ such that $R_2(I) > R_1(I)$ for all $I$, but an entrepreneur would prefer to undertake the “worse” project $R_1(\cdot)$ because doing so delivers higher welfare.

Formally, a project is linear-flat if the return function can be parameterized by a slope parameter $\alpha$ and a cutoff parameter $L$ such that $R(I; \alpha, L) = \alpha \min(I, L)$. Define $L^{SL}(\alpha)$ and $D^{SL}(\alpha)$ respectively as the market and face value of debt that would be chosen by an entrepreneur who can commit to borrowing from a single lender when the investment opportunity is linear with marginal returns $\alpha$. Now consider the linear-then-flat project $R(I; \alpha, L^{SL}(\alpha))$, which is linear with slope $\alpha$ and turns flat after raising an investment level of $L^{SL}(\alpha)$. It should be clear that even without commitment power this project achieves the same lending outcome as the linear project with slope $\alpha$ under commitment: as there are no marginal returns from borrowing beyond $(D^{SL}(\alpha), L^{SL}(\alpha))$, the first lender that the entrepreneur visits would be willing to accept a loan at these terms since it will be assured that no future borrowing will occur to deteriorate the value of the claims. In this example the extreme concavity of investment opportunity effectively solves the commitment problem. We now use this result to illustrate Proposition 3.7.

Figure 3-6 shows the payoff functions of several projects. Consider Project 1 to be the linear-flat project with slope $\alpha^L$ and a threshold for zero marginal return at $L^{SL}(\alpha^L)$, i.e. $R_1(I) = \alpha^L \min(I, L^{SL}(\alpha^L))$. For a given $q > 0$, we can find a fully linear Project 2 with slope $\alpha^H$ such that the equilibrium welfare obtained by an entrepreneur endowed with either of these two projects would be equal. We must have $\alpha^H > \alpha^L$ because the project needs to deliver higher output to compensate for the surplus loss due to the lack of commitment power.

Any point in the graph can be used to parameterize a project payoff function that is linear from the origin to that point, then turns flat after that. The dotted line plots the set of points that
represent projects which give equivalent welfare to the entrepreneur as Project 1 and Project 2. Any project in the shaded region, such as Project 3, gives strictly higher welfare in equilibrium to the entrepreneur than Project 1 does. Any project outside the shaded region and below Project 2 gives strictly lower welfare.

The strength of this result can be seen by comparing Project 3 and Project 2. Despite the fact that viewed in isolation, Project 3 yields lower return than Project 2 for any level of investment, it gives the entrepreneur strictly higher welfare in our model due to the commitment friction. By having zero marginal return beyond a certain level of investment, Project 3 endows the entrepreneur with some endogenous commitment power such that loans can be obtained under better interest rates. As a result, if the entrepreneur gets to choose ex-ante which investment projects to undertake, he may endogenously pursue an investment opportunity with lower returns, as long as it is sufficiently concave.

The result in this section highlights that in the presence of the commitment problem there is a powerful trade-off between the average and marginal returns of investment opportunities. Productivity is no longer the sole determinant of investment in economic activities. The commitment friction can be mitigated by undertaking instead in projects that have concave returns and thus embed some degree of commitment. Once these projects interior to the technology frontier are undertaken, they exhibit slower capital accumulation and growth potential.
3.6 Policy Implications

We now turn to studying simple regulatory policy tools that can improve outcomes in our model: limiting interest rates, imposing total borrowing limits, and limiting the number of lenders from which a borrower can obtain loans. Conventional arguments suggest that interest rate caps may be helpful in improving allocations when there is a lack of competition among lenders as they limit monopoly power. Yet in some scenarios, it seems to have been the entry of new lenders into markets and the resulting increase in competition that has driven regulators to consider usury regulations. The microfinance crisis in Andhra Pradesh was precipitated by the rapid entry of thousands of new microfinance lenders and characterized by over indebtedness of borrowers from multiple lenders. A report commissioned by the Reserve Bank of India to study the causes and potential regulatory responses to the Andhra Pradesh crisis states:

"It has been suggested that with the development of active competition between MFIs there has been a deluge of loan funds available to borrowers which has fueled excessive borrowing and the emergence of undesirable practices ... Finally, it is believed that in consequence of over-borrowing, default rates have been climbing in some locations but these have not been disclosed because of ever-greening and multiple lending." Malegam (2011).

Despite highlighting a high degree of competition, the report proceeds to propose regulation that limits interest rates charged to borrowers. The following proposition illustrates that, through the lens of our model of multiple borrowing and commitment problems, this type of regulatory response is rational and welfare improving.

**Proposition 3.8.** Adding an upper bound on interest rates generates the following results:

a) For any increasing and concave investment opportunity \( R(\cdot) \) with \( R'(0) > 1 \), there is an optimal interest rate cap \( \bar{r}_{SL}(R) \) that induces the full commitment allocation with the borrower obtaining funding from a single lender. When \( R(I) = \alpha I \), the optimal interest rate cap is \( \bar{r}_{SL} \equiv 1 - \alpha^{-1} \).

b) Any interest rate cap \( \bar{r} \) has an associated debt limit \( \bar{D} = 1 - (1 + \bar{r})^{-1} \) that induces the same equilibrium.

c) If \( \bar{r} < \bar{r}_{SL} \) then single lender borrowing prevails but debt is inefficiently low and the interest rate cap is too restrictive. Welfare will be lower than the unregulated equilibrium (i.e. with no interest rate cap) if the interest rate cap is sufficiently low.

d) If \( \bar{r} > \bar{r}_{SL} \) then the interest rate cap is too loose. Imposing the cap increases expected investment and welfare while lowering expected debt and interest rates relative to the unregulated equilibrium. As \( \bar{r} \to \infty \) the interest rate cap becomes irrelevant and outcomes converge to the unregulated equilibrium.
Interest rate caps have the potential to improve welfare because they embed commitment power. Recall that in the model the more debt the borrower has outstanding, the lower the probability of repayment, and thus the higher interest rates need to be on additional lending. Interest rate caps add commitment power because marginal borrowing at high enough levels of debt would need to violate the interest rate cap for these loans to break even and thus are never issued. Early lenders can then be assured that such future borrowing, which increases the anticipated probability of default, will not occur, and can provide initial loans at interest rates that are closer to those that would prevail if borrowers could fully commit to exclusive borrowing. This better aligns borrower incentives, reducing face values of debt and interest rates in equilibrium. Importantly, interest rate caps also increase investment and improve welfare as long as they are not too severe that they prevent productive investment.

Further, for a given investment opportunity the interest rate cap can be set to fully overcome the inefficiencies induced by lack of commitment. By setting the interest rate cap at exactly the interest rate that would prevail in the full-commitment equilibrium, the borrower can credibly raise exactly the full commitment level of debt, and at the commitment-level interest rate, restoring the full commitment outcomes. By proposing such a loan to the first lender, the lender is assured that any future borrowing would necessarily need to be at interest rates above the allowed limit, and can thus be sure that no such additional borrowing would take place. Thus this loan proposal is expected to earn zero profits for the lender and is always accepted. By definition this allocation maximizes the entrepreneur’s welfare given lenders at least break even, so any rational lender strategy induces this loan proposal as a best response.

In our model a limit on the total face value of debt a borrower can obtain is mechanically equivalent to a particular interest rate cap, and the equivalent debt limit is increasing in the interest rate cap, so all the results above also apply to total debt limits. Both total borrowing limits and interest rate caps were adopted for microfinance loans in India in 2011 (Dr. D. Subbarao (2011)). These policies are equivalent in the model because they both operate by shutting down contingencies of excessive debt accumulation: debt limits directly and interest rate limits indirectly through the fact that the lower bound on interest rates in any equilibrium is increasing in the total cumulative face value of debt. In the contingencies (potentially off-equilibrium) in which either the interest rate or debt limit is binding, the lender zero profit condition implies a unique relationship between interest rates and total face value of debt at this point: \( p(D) = (1 + \bar{r})^{-1} \), where \( D \) is a borrowing limit and \( \bar{r} \) is the equivalent interest rate limit.

It is important to re-emphasize that in general interest rate caps (and total borrowing limits) have ambiguous implications for welfare, because caps that are too low can restrict productive investment. While there are always welfare improving interest rate limits for a given project, imposing a market-wide policy can have ambiguous effects on welfare if the investment opportunities in the economy are sufficiently heterogeneous. From a utilitarian perspective, however, “reasonable” interest rate caps can be quite beneficial if many projects in the economy could benefit from them.
Further, taking into account the possibility of endogenous project choice, interest rate caps have the potential to "unlock" the best projects available that were previously infeasible due to the endogenous credit constraint induced by lack of commitment.

Finally, a surprising and controversial policy recommendation of Malegam (2011) was to limit borrowers to obtaining loans from at most two micofinance lenders. Limiting the number of lenders from which a borrower can obtain loans will mechanically increase commitment power by limiting the opportunities a borrower has not to commit. While this policy was not ultimately adopted, our model shows that in the face of commitment problems such a policy may actually be quite helpful.

3.7 Conclusion

This paper argues that commitment problems in lending markets can explain emerging empirical evidence that the rapid expansion of credit access can have perverse effects. When borrowers cannot commit to exclusive contracting, increasing the availability of lenders makes markets appear less competitive as interest rates rise and entrepreneur investment and welfare fall. More importantly, commitment problems can result in better projects receiving less investment than worse projects. This force can be so severe that what look like good opportunities are passed over for inferior investment technology. Finally, we show how simple regulatory tools such as interest rate ceilings and debt limits can improve outcomes and ameliorate the misallocative forces we highlight.

The intuition for these result is that the externalities the lenders impose on each other when commitment or contingent contracting is not possible can prevent the borrower from being able to use pledgeable cash flows efficiently. When explicit commitment is impossible, there is value in any implicitly commitment mechanisms that attenuate the demand for further borrowing. Thus, the return profile of an investment opportunities itself is an important driver in the severity of commitment distortions.

Since commitment is less of a problem for projects with lower marginal returns, when given the choice entrepreneurs will endogenously choose investment opportunities that are everywhere less productive than other available opportunities, as long as they are sufficiently more concave. Thus our model provides a new micro-foundation for the idea that commitment problems in lending markets can induce substantial misallocation in capital investment and can explain observations both of low growth and of economic activity below the technological frontier. While we have augmented our study with a sample of the growing anecdotal evidence that multiple borrowing is problematic, there is much more to learn. Formally testing the empirical validity the mechanisms we highlight in explaining the failure of increased access to finance to significantly improve outcomes is an important topic for future research.
Appendix

Microfounding Random Default with Moral Hazard

In this section we provide an alternative microfoundation of the random default through limited pledgeability and moral hazard. Instead of assuming the entrepreneur borrows against his default cost $\bar{c}$ as we do in the main text of the paper, we endow the entrepreneur with a pledgeable stochastic cashflows from assets-in-place that are realized at $s = 2$. The pledgeable cashflows realize one of two values: zero or one. The good payoff occurs with probability $p$, which is chosen by the entrepreneur at quadratic effort cost $p^2$. Moral hazard arises because effort is chosen after cashflows are (partially) pledged to lenders. Since there are only two realizations of these cashflows and the entrepreneur has no other pledgeable wealth, it is without loss of generality that claims issued to lenders take the form of debt contracts: they are repaid in full when the good realization occurs and the entrepreneur defaults when cashflows are zero. With this in mind, assume the entrepreneur has issued a total face value of debt $D$ in the first stage. At $s = 2$ the entrepreneur solves

$$p (D) \equiv \arg \max_p \{1 - D\} - p^2$$

The entrepreneur expects the assets in place to pay out with probability $p$, and conditional on a positive payout the entrepreneur gets to keep $1 - D$ of the cashflows as the residual claim. The cost of choosing the probability of positive cashflows to be $p$ is $p^2$. Thus $p (D)$ denotes the solution to the entrepreneur’s choice of effort at stage $s = 2$ conditional on having a total outstanding face value of debt $D$. The solution of entrepreneur’s problem yields $p (D) = \frac{1-D}{2}$, and entrepreneur’s expected payoff from the residual claim is $\frac{(1-D)^2}{4}$ The value function in (3.3.4) is modified to

$$V (D) = \max_{D'} \alpha \hat{p} (D') (D' - D) - (1 - q) \frac{(1-D)^2}{4} + qV (D')$$

and all of our results go through analogously.

Proof of Lemma 3.1

Borrower’s problem can be formulated recursively as:

$$V (D) = \max_{D'} \alpha \hat{p} (D') (D' - D) + (1 - q) \left[ \frac{(1-D')^2}{2} - \frac{1}{2} \right] + qV (D')$$

where

$$\hat{p} (D') \equiv \mathbb{E} [1 - D^{agg} | D']$$

228
We guess that borrower’s policy function \( g(D) \) and lenders’ loan pricing function \( \tilde{p}(\cdot) \) both take a linear form and are each characterized by a single endogenous variable, \( b \) and \( \ell \), respectively:

\[
\tilde{p}(D) = \ell^{-1}(1-D) \\
1 - g(D) = b^{-1}(1-D)
\]

To solve for borrower’s policy function, we proceed to take first order condition and use the envelope condition for borrower’s problem. The first order condition is:

\[
-\alpha \ell^{-1}(g(D) - D) + \alpha \ell^{-1}(1 - g(D)) - (1 - q)(1 - g(D)) + qV'(g(D)) = 0
\]

and the envelope condition is:

\[
V'(D) = -\alpha \ell^{-1}(1 - g(D))
\]

Plugging the envelope condition into the first-order condition and after simplifying, we can express the Euler condition as a quadratic function of \( b^{-1} \):

\[
q\alpha \ell^{-1}b^{-2} + (1 - q - 2\alpha \ell^{-1})b^{-1} + \alpha \ell^{-1} = 0
\]

We thus solve for the endogenous parameter \( b^{-1} \) that governs the borrower’s policy function as:

\[
b^{-1} = \frac{(2\alpha \ell^{-1} - (1 - q)) - \sqrt{(1 - q - 2\alpha \ell^{-1})^2 - 4q\alpha^2\ell^{-2}}}{2q\alpha \ell^{-1}} \tag{3.7.1}
\]

To solve for the lender’s loan pricing function, note

\[
\tilde{p}(D) = \mathbb{E}[1 - D^{agg}|D] \\
= (1 - q) \left[ (1 - D) + q (1 - g(D)) + q^2 (1 - g(g(D))) + \cdots \right] \\
= (1 - q) \left[ (1 - D) + qb^{-1}(1 - D) + q^2b^{-2}(1 - D) + \cdots \right] \\
= \frac{1 - q}{1 - qb^{-1}} (1 - D)
\]

Hence

\[
\ell^{-1} = \frac{1 - q}{1 - qb^{-1}} \tag{3.7.2}
\]

Equation (3.7.1) characterizes \( b^{-1} \) as a decreasing function of \( \ell^{-1} \). On the other hand, equation (3.7.2) characterizes \( \ell^{-1} \) as an increasing function of \( b^{-1} \). The two equations therefore yields a

---

\(^{10}\)There are two roots to the quadratic equation, one of which leads to explosive debt accumulation. We choose the other, stable root.
unique solution \((\bar{b}^*, \ell^*)\) for each \(q \in [0, 1)\) and \(\alpha \in (1, \infty)\). In particular, we have

\[
(\bar{b}^*)^{-1} = \frac{2\alpha - 1 - \sqrt{4(1-q)(\alpha^2 - \alpha) + 1}}{2q(\alpha - 1)}
\]

Lemma 3.2. The best response functions have the following properties:

\[
\frac{\partial \ell (b; \alpha, q)}{\partial b} \geq 0; \quad \frac{\partial \ell (b; \alpha, q)}{\partial \alpha} = 0;
\]

\[
\frac{\partial b (\ell; \alpha, q)}{\partial \alpha} \geq 0; \quad \frac{\partial b (\ell; \alpha, q)}{\partial \ell} \leq 0;
\]

where the inequalities are strict for \(q \in (0, 1)\) and \(\alpha > 1\).

Proof. The results with respect to lender’s best response immediately follow from equation (3.7.2).

We now with with equation (3.7.1) to derive the results respect to borrower’s best response function. Let \(x = 2\alpha \ell^{-1}\), we have

\[
b^{-1} = \frac{(x - (1-q)) - \sqrt{(1-q) - x^2 - qx^2}}{qx}
\]

\[
= \frac{1}{q} \left( 1 - q + \left( (1-q)(x-1)^2 - q(1-q) \right)^{\frac{1}{2}} \right)
\]

Let \(\Delta \equiv \left( (1-q)(x-1)^2 - q(1-q) \right)\) and take derivative with respect to \(x\), we have

\[
\frac{\partial b^{-1}}{\partial x} = -\frac{qx(1-q)(x-1) - q\left( (1-q)\Delta^{\frac{1}{2}} + (1-q)(x-1)^2 - q(1-q) \right)}{(qx)^2 \Delta^{\frac{1}{2}}}
\]

\[
= -\frac{q(1-q)}{(qx)^2 \Delta^{\frac{1}{2}}} \left( x^2 - x - \Delta^{\frac{1}{2}} - (x-1)^2 + q \right)
\]

\[
= -\frac{q(1-q)}{(qx)^2 \Delta^{\frac{1}{2}}} \left( x + q - 1 - \Delta^{\frac{1}{2}} \right)
\]

Since \(\Delta^{\frac{1}{2}} = \sqrt{(x+q-1)^2 - qx^2} \leq x + q - 1\), we have that

\[
\frac{\partial b^{-1}}{\partial x} \leq 0
\]

and the inequality is strict for \(q \in (0, 1)\). Given the definition of \(x\), we have that \(\frac{\partial b (\ell; \alpha, q)}{\partial \alpha} \geq 0\) and \(\frac{\partial b (\ell; \alpha, q)}{\partial \ell} \leq 0\). \(\Box\)
Proof of Proposition 3.3

We begin with an outline of the proof. The proof will first show how to solve the finite-lender game by backwards induction, generating a recursive formulation for borrower and lender strategies. Next, we show that the fixed point of this recursion generates the strategies of the infinite-lender equilibrium defined in the main text. Finally, to demonstrate convergence, we show that this recursive formulation of strategies is characterized by a contraction mapping. This implies that, considering the strategies at a given lender, as the number of potential subsequent lenders goes to infinity, the equilibrium strategies at this lender converge uniquely to the fixed point and thus to the strategies of the infinite-lender equilibrium.

Backward Induction in the Finite-Lender Game

Consider a finite version of the game with $N$ lenders. For this proof, we abuse notation and index periods counting backwards from the end. Thus the last lender is indexed 1, and the first lender is indexed $N$. Therefore, after lender $i$ there are at most $i - 1$ more lenders for the borrower to visit. Let $D_i$ denote the amount of cumulative debt the borrower accumulates from meeting lenders $N$ through $i + 1$, thus $D_0$ denotes the total amount of debt the borrower will accumulate if it gets to meet all $N$ lenders. The probability of default from the last lender’s perspective will be $\tilde{p}_1(D_0) = 1 - D_0$ if the lender accepts a proposal that brings the borrower’s cumulative debt to $D_0$. This defines the unique strategy of the final lender to accept only weakly profitable loans. Assume the borrower has any arbitrary face value of debt $D_1$ upon meeting the final lender. The borrower solves

$$V_1(D_1) = \max_{D_0} \alpha (D_0 - D_1)(1 - D_0) + \left(\frac{D_0^2}{2} - D_0\right)$$

and the solution is

$$1 - D_0 = (1 - D_1) \frac{\alpha}{2\alpha - 1}$$

with borrower’s maximized value function being

$$V_1(D_1) = (1 - D_1)^2 \left[\frac{2\alpha^3 - \alpha^2}{2(2\alpha - 1)^2}\right] - \frac{1}{2}$$

We define

$$B_1 = \frac{\alpha}{2\alpha - 1}$$
$$L_1 = 1$$
$$W_1 = \frac{2\alpha^3 - \alpha^2}{2(2\alpha - 1)^2}$$
Thus the unique subgame perfect equilibrium strategies conditional on arriving to the last lender with some amount of debt $D$ are:

\[
1 - g(D) = B_1 (1 - D) \\
\bar{p}_1(D) = L_1 (1 - D) 
\]

and any borrower considering leaving the second-to-last lender with a total face value of debt $D$ realizes that continuation utility if it reaches the last lender is given by

\[
V_1(D) = W_1 (1 - D)^2 - \frac{1}{2} 
\]

Thus we know that at lender $i = 1$ players use strategies linear in $1 - D$. Now we show by induction that all lenders use such linear strategies. Assume for some $n$ that players at all stages $i < n$ use linear strategies and that the maximized value function at lender $i$ is proportional to $(1 - D_{i+1})^2$. We will show that players at stage $n$ also use linear strategies and that the maximized value function at lender $n$ is proportional to $(1 - D_{n+1})^2$, and thus by induction prove that these claims do indeed hold for all $n \in \mathbb{N}$.

Now consider the subgame where the borrower meets lender $n$ with cumulative debt $D_n$ obtained from previous lenders. Since all future lenders and borrowers use linear strategies, we can compute lender $n$'s expected probability of repayment:

\[
\bar{p}_n(D_{n-1}) = \left(1 - q\right)(1 - D_{n-1}) + q\bar{p}_{n-1}(D_{n-2}) \\
= \left(1 - q\right)(1 - D_{n-1}) + qL_{n-1} (1 - D_{n-2}) \\
= \left(1 - q\right)(1 - D_{n-1}) + qL_{n-1}B_{n-1} (1 - D_{n-1}) \\
= \left[\left(1 - q\right) + qB_{n-1}L_{n-1}\right] (1 - D_{n-1}) 
\]

where the first to second line follows from the assumption that lender $n - 1$ is using a linear strategy $\bar{p}_{n-1}(D) = L_{n-1} (1 - D)$, and moving from the second to the third line relies on the assumption that the borrower at $n - 1$ is using a linear strategy $1 - D_{n-2} = B_{n-1} (1 - D_{n-1})$. Thus we know that lender $n$ follows the strategy given by

\[
L_n = 1 - q + qB_{n-1}L_{n-1} 
\]

Next, under our inductive hypothesis we can write the borrower's problem visiting lender $n$ as:
\[ V_n (D_n) = \max_{B_{n-1}} \alpha B_{n-1} (D_{n-1} - D_n) + (1 - q) \left( \frac{D_{n-1}^2}{2} - D_{n-1} \right) + q V_{n-1} (D_{n-1}) \]

\[ = \max_{B_{n-1}} \left\{ \alpha (D_{n-1} - D_n) (1 - D_{n-1}) L_n + (1 - q) \left( \frac{D_{n-1}^2}{2} - D_{n-1} \right) \right. \]

\[ + q \left( W_{n-1} (1 - D_{n-1})^2 - \frac{1}{2} \right) \right\} \]

Taking the first order condition and solving or \( D_{n-1} \) verifies that the borrower's strategy does indeed have the hypothesized linear form:

\[ 1 - D_{n-1} = (1 - D_n) \frac{\alpha L_n}{2 \alpha L_n - (1 - q + 2q W_{n-1})} \]

and maximized value function

\[ V_n (D_n) = (1 - D_n) \left( \alpha (1 - B_n) B_n L_n + (1 - q) \frac{B^2_n}{2} + q B^2 W_{n-1} \right) \]

\[ = W_n \]

Thus the inductive proof is completed and all strategies satisfy the proposed form.

**Recursive Formulation of Strategies**

It is clear from above that the vector \((B_n, L_n, W_n)\) is generated by a system of 3 difference equations:

\[ L_n = (1 - q) + q B_{n-1} L_{n-1} \]

\[ B_n = \frac{\alpha L_n}{2 \alpha L_n - (1 - q + 2q W_{n-1})} \]

\[ W_n = \alpha (1 - B_n) B_n L_n + (1 - q) \frac{B^2_n}{2} + q B^2 W_{n-1} \]

rearranging so each of \((L_n, B_n, W_n)\) is a function of only the lagged variables:

\[ L_n = 1 - q + q B_{n-1} L_{n-1} \]

\[ B_n = \frac{\alpha (1 - q + q B_{n-1} L_{n-1})}{2 \alpha (1 - q + q B_{n-1} L_{n-1}) - (1 - q + 2q W_{n-1})} \]

\[ W_n = \frac{\alpha^2 (1 - q + q B_{n-1} L_{n-1})^2}{4 \alpha (1 - q + q B_{n-1} L_{n-1}) - 2 (1 - q + 2q W_{n-1})} \]
where the last equation can be simplified to

\[ W_n = \frac{\alpha}{2} B_n L_n. \]

**Convergence to Infinite-Lender Strategies**

Defining a new variable \( x_n \equiv B_n L_n \), the set of difference equations above can be rewritten as

\[
L_n = (1 - q) + q x_{n-1} \\
B_n = \frac{\alpha ((1 - q) + q x_{n-1})}{2 \alpha ((1 - q) + q x_{n-1}) - (1 - q + \alpha q x_{n-1})} \\
W_n = \frac{\alpha}{2} x_n
\]

Hence the sequence \( \{x_n\} \) is defined by \( x_1 \equiv B_1 L_1 = \frac{\alpha}{2 \alpha - 1} \) and a continuous function \( f(\cdot) \) such that \( x_n = f(x_{n-1}) \), where

\[
f(x) = \frac{\alpha ((1 - q) + qx)^2}{(1 - q)(2\alpha - 1) + \alpha qx}
\]

First note that \( f(x) > 0 \) if \( x > 0 \), and since \( x_1 > 0 \), we have \( x_n > 0 \) for all \( n \). It is also easily verified that \( f(\cdot) \) has a unique fixed point \( x^* \), which corresponds to the unique fixed point \( (B^*, L^*, W^*) \) of the system of difference equations above. Moreover, it is a matter of algebra to show that the fixed point coincides with our closed form solution of the infinite lender game, implying that \( b^{*-1} = B^* \) and \( \ell^{*-1} = L^* \).

We next show that \( f(\cdot) \) defines a contraction mapping which, given the continuity of \( f(\cdot) \), shows that \( x_n \rightarrow x^* \). This in turn implies \( (B_n \rightarrow b^{*-1}, L_n \rightarrow \ell^{*-1}) \) as \( n \rightarrow \infty \). In words, this means that as the number of potential future lenders in the game after lender \( n \) grows towards infinity, the unique subgame perfect strategies the players at stage \( n \) converge to the stationary strategies employed by all players in the infinite-lender game.

To show \( f(\cdot) \) defines a contraction mapping, we show \( ||f'(x)|| < 1 \). Taking derivatives of \( f \) with respect to \( x \), we have (after much simplification)

\[
f'(x) = q - \frac{q (\alpha - 1)^2 (1 - q)^2}{(2\alpha + q - 2\alpha q + \alpha qx - 1)^2} \\
= q \left( 1 - \left( \frac{(\alpha - 1)(1 - q)}{(2\alpha - 1)(1 - q) + \alpha qx} \right)^2 \right)
\]
Since \( x > 0, 1 > q \geq 0, \alpha > 1 \), we have

\[
0 < \frac{(\alpha - 1)(1 - q)}{(2\alpha - 1)(1 - q) + \alpha qx} < 1
\]

\[
0 < \frac{(\alpha - 1)(1 - q)}{(\alpha - 1)(1 - q) + \alpha (1 - q) + \alpha qx} < 1
\]

Hence

\[
0 < f'(x) < 1
\]

which shows that \( f(\cdot) \) is a contraction mapping.

\[\square\]

**Proof of Proposition 3.4**

Before proving the comparative static propositions, it will be useful to derive some expressions for equilibrium objects of interest in terms of parameters \( \alpha \) and \( q \). Let \( I^{agg} \) denote the aggregate investment and \( D^{agg} \) denote the aggregate debt that have been attained when the lending market game ends. In equilibrium, ex-ante, these are random variables with respect to the number of lenders the borrower will be able to visit.

**Lemma 3.3.** \( \mathbb{E}[I^{agg}] = \mathbb{E}[p(D^{agg})D^{agg}] \)

*Proof.* Let \( N \) denote the random number of lenders the borrower gets to visit before losing access to the lending market game. The random aggregate face value of debt and aggregate investment can be expressed as:

\[
D^{agg} = \sum_{j=1}^{\infty} D_j 1(N \geq j)
\]

\[
I^{agg} = \sum_{j=1}^{\infty} I_j 1(N \geq j)
\]

where \( D_j \) is the amount of debt given by the \( j \)-th lender. Similarly denote \( I_j \) to be the amount of investment capital provided by the \( j \)-th lender. Pick any \( j > 0 \), the zero-profit condition for his loan and investment size is:

\[
\mathbb{E}[p(D^{agg}) | N \geq j] D_j 1(N \geq j) = I_j 1(N \geq j)
\]

Taking expectation over \( N \) on both sides and applying the law of iterated expectation, we get:

\[
\mathbb{E}[p(D^{agg}) D_j 1(N \geq j)] = \mathbb{E}[I_j 1(N \geq j)]
\]

We next sum the previous equation over all lenders. By the linearity of the expectations operator,
we can bring the sum inside:
\[
\mathbb{E}
\left[
p(D^{agg}) \sum_{j=1}^{\infty} D_j \mathbf{1}(N \geq j)
\right]
= \mathbb{E}
\left[
\sum_{j=1}^{\infty} I_j \mathbf{1}(N \geq j)
\right]
\]

Substituting in the definitions of \(D^{agg}\) and \(I^{agg}\):
\[
\mathbb{E}[p(D^{agg}) D^{agg}] = \mathbb{E}[I^{agg}]
\]

Lemma 3.4. We can express the expected debt and investment as functions of \(b^{-1}\):
\[
\mathbb{E}[D^{agg}] = \frac{1 - b^{-1}}{1 - q b^{-1}}
\]
\[
\mathbb{E}[I^{agg}] = \frac{b^{-1} (1 - b^{-1}) (1 - q)}{(1 - b^{-1} q) (1 - b^{-2} q)}
\]

Proof. Denote the expected aggregate debt upon leaving a given lender with cumulative debt \(D\) as \(\mathbb{E}[D^{agg}|D]\). From lender's zero-profit condition, we have
\[
\mathbb{E}[D^{agg}|D] = 1 - \bar{p}(D)
\]

The ex-ante expected aggregate debt \(\mathbb{E}[D^{agg}]\) is simply the expected aggregate debt upon leaving a lender with zero outstanding debt, times \(\frac{1}{q}\) (since the borrower meets the first lender with certainty, not probability \(q\)). Thus we have
\[
\mathbb{E}[D^{agg}] = \frac{1}{q} \mathbb{E}[D^{agg}|0]
= \frac{1}{q} \left( 1 - \frac{1 - q}{1 - q b^{-1}} \right)
= \frac{1 - b^{-1}}{1 - q b^{-1}}
\]
To get the expression for expected investment:

\[ E[I] = \tilde{p}(g(0)) g(0) + q\tilde{p}\left(g^2(0)\right) \left[g^2(0) - g(0)\right] + q^2\tilde{p}\left(g^3(0)\right) \left[g^3(0) - g^2(0)\right] + \ldots \]

\[ = \epsilon^{-1}\left[(1 - g(0)) g(0) + q (1 - g^2(0)) \left[(1 - g(0)) - (1 - g^2(0))\right] + \ldots\right] \]

\[ = \epsilon^{-1}\left[b^{-1} (1 - b^{-1}) + qb^{-2} \left[b^{-1} - b^{-2}\right] + q^2b^{-3} \left[b^{-2} - b^{-3}\right] + \ldots\right] \]

\[ = \epsilon^{-1}b^{-1} \left(1 - b^{-1}\right) \left[1 + qb^{-2} + q^2b^{-4} + \ldots\right] \]

\[ = \frac{b^{-1} \left(1 - b^{-1}\right) (1 - q)}{(1 - b^{-1}q) (1 - qb^{-2})} \]

\[ \square \]

**Lemma 3.5.** Let \( z = \sqrt{4(1 - q)(\alpha^2 - \alpha) + 1} \). The analytic solution of expected debt, investment, and welfare can be expressed as the following functions of parameters \( q \) and \( \alpha \):

\[ E[D^{agg}] = \frac{2\alpha - 1 - z}{2q\alpha} \]

\[ E[I^{agg}] = \frac{(\alpha - 1) (z + 1 - 2\alpha (1 - q))}{2q(2\alpha - 1)} \]

\[ V(0) = \frac{1 - 2\alpha (1 - q) - 2q + z}{4q} \]

*Proof.* These expressions can be obtained by substituting the analytic solution of \( b^* \) from lemma 1 into the expressions in lemma 2. \( \square \)

Now continuing on, we can express equilibrium \( b^* \) as

\[ (b^*)^{-1} = \frac{2\alpha - 1 - z}{2q(\alpha - 1)} \]

Also note that

\[ \frac{\partial z}{\partial \alpha} = z^{-1} (1 - q) (4\alpha - 2) \]

\[ \frac{\partial z}{\partial q} = -2z^{-1} (\alpha^2 - \alpha) \]

We now proceed to prove Proposition 3.4 claim by claim.

*Claim 1.* \( E[D^{agg}] \) is decreasing in \( q \).
Proof. We first express $\mathbb{E}[D^{agg}]$ as a function of $\alpha$, $q$, and $z$:

$$
\mathbb{E}[D^{agg}] = \frac{1 - \frac{2^{\alpha-1}z}{2q^{\alpha-1}}}{1 - \frac{2^{\alpha-1}z}{2^{(\alpha-1)}}} = \frac{2q(\alpha-1) - 2\alpha + 1 + z}{2q(\alpha-1) - 2q\alpha + q + qz}
$$

Therefore:

$$
\mathbb{E}[D^{agg}] = \frac{(2q\alpha - 2q - 2\alpha + 1 + z)(z + 1)}{q(z - 1)(z + 1)} = \frac{(2\alpha - 1 - z)(1 - q)(\alpha - 1)}{2q\alpha(1 - q)(\alpha - 1)} = \frac{2\alpha - 1 - z}{2q\alpha}
$$

Differentiating with respect to $q$, we get

$$
\frac{d\mathbb{E}[D^{agg}]}{dq} = \frac{-2q\alpha}{(2q\alpha)^2} \frac{dz}{dq} - \frac{(2\alpha - 1 - z)2\alpha}{(2q\alpha)^2}
$$

which implies

$$
sign\left(\frac{\partial \mathbb{E}[D^{agg}]}{\partial q}\right) = sign\left(2q\alpha^2 - \alpha - (2\alpha - 1 - z)z\right) = sign\left(2q\alpha^2 - \alpha + 4(1 - q)\alpha^2 + 1 - (2\alpha - 1)z\right) = sign\left((\alpha^2 - \alpha)(4 - 2q) + 1 - (2\alpha - 1)z\right)
$$

Let $RHS \equiv (\alpha^2 - \alpha)(4 - 2q) + 1 - (2\alpha - 1)z$. The remaining proof consists of three steps: 1) show $\frac{dRHS}{d\alpha} \geq 0$ for all $\alpha \geq 1$, $q \in [0, 1]$; 2) show $\frac{dRHS}{dq} \geq 0$ for all $\alpha \geq 1$, $q \in [0, 1]$, with equality holding only when $\alpha = 1$ or $q = 0$; 3) RHS evaluated at $\alpha = 1$, $q = 0$ is zero, concluding that $RHS > 0$ for $\alpha > 1, q > 0$. 

238
Step 1: show $\frac{dRHS}{d\alpha} > 0$ for all $\alpha \geq 1$, $q \in [0,1]$. Differentiating $RHS$ with respect to $\alpha$, we have

$$
\frac{dRHS}{d\alpha} = (2\alpha - 1) (4 - 2q) - 2z - (2\alpha - 1) \frac{\partial z}{\partial \alpha} \\
= (2\alpha - 1) \left( 4 - 2q - z^{-1} (1 - q) (4\alpha - 2) \right) - 2z \\
= (2\alpha - 1) \left( 4 - 2q - z^{-1} (1 - q) (4\alpha - 2) \right) \\
\geq (2\alpha - 1) \left( 4 - \max_{q\in[0,1]} (2q) - \max_{q\in[0,1]} z^{-1} (1 - q) (4\alpha - 2) \right) \\
= (2\alpha - 1) (4 - 2 - 2) \\
= 0
$$

Step 2: show $\frac{dRHS}{dq} > 0$ for all $\alpha \geq 1$, $q \in [0,1]$, with equality holding only when $\alpha = 1$ or $q = 0$. Differentiating $RHS$ with respect to $q$, we have

$$
\frac{dRHS}{dq} = -2 \left( \alpha^2 - \alpha \right) - (2\alpha - 1) \frac{\partial z}{\partial q} \\
= \left( \alpha^2 - \alpha \right) \left( 2z^{-1} (2\alpha - 1) - 2 \right) \\
= 2z^{-1} \left( \alpha^2 - \alpha \right) \underbrace{(2\alpha - 1 - z)}_{\geq 0} \\
> 0
$$

The last term is non-negative and is zero only when $q = 0$. To see this, note

$$
z = \sqrt{4 (1 - q) (\alpha^2 - \alpha) + 1} \\
\text{(equal only if } q = 0) \leq \sqrt{4\alpha^2 - 4\alpha + 1} \\
= 2\alpha - 1
$$

Hence we have $\frac{dRHS}{dq} > 0$.

Step 3: conclude the proof. Note

$$RHS|_{q=0,\alpha=1} = 0$$

Hence we have, for any $\alpha > 1$ and $q > 0$, $RHS > 0$. Thus $\frac{\partial E[I^{agg}]}{\partial q} > 0$. \qed

Claim 2. $E[I^{agg}]$ is decreasing in $q$.

Proof. Given the result in lemma (3.5), we first show that investment is decreasing in $q$ if and only

$$
if the exante welfare for the borrower is decreasing in $q$. To see this, note

$$
\mathbb{E} [T^{agg}] = \frac{(\alpha - 1) (z + 1 - 2\alpha (1 - q))}{2\alpha q (2\alpha - 1)}
$$

$$
= \frac{2 (\alpha - 1)}{\alpha (2\alpha - 1)} \left[ \frac{(z + 1 - 2\alpha (1 - q) - 2q)}{4q} + \frac{1}{2} \right]
$$

$$
= \frac{2 (\alpha - 1)}{\alpha (2\alpha - 1)} \left[ V(0) + \frac{1}{2} \right]
$$

To show $V(0)$ is decreasing in $q$, first note

$$
V(0) = \frac{1 - 2\alpha (1 - q) - 2q + z}{4q}
$$

$$
\frac{dV(0)}{dq} = \frac{4q (2\alpha - 2 + qz) - 4 (1 + 2\alpha (q - 1) - 2q + z)}{16q^2}
$$

$$
= \frac{1}{4q^2} [q (2\alpha - 2 + qz) - 1 + 2\alpha (1 - q) + 2q - z]
$$

$$
= \frac{1}{4q^2} [-2\alpha \left( (\alpha - 1) \left( \frac{q}{z} \right) - 1 \right) - (1 + z)]
$$

Define $\Omega = -2\alpha \left[ (\alpha - 1) \left( \frac{q}{z} \right) - 1 \right] - (1 + z)$, we then have $\text{sign} \left( \frac{dV(0)}{dq} \right) = \text{sign} (\Omega)$.

To compute the derivative of $\Omega$ with respect to $q$:

$$
\frac{d\Omega}{dq} = -2\alpha (\alpha - 1) \frac{z - qzq}{z^2} - zq
$$

$$
= -\frac{1}{z^3} 4\alpha^2 (\alpha - 1)^2 q \leq 0
$$

Therefore $\Omega$ is declining in $q$ for all $\alpha > 1$. This means to check that $\frac{dV(0)}{dq} < 0$ it is sufficient to check that $\Omega|_{q=0} \leq 0$.

$$
\Omega (q = 0) = (2\alpha - 1) - z
$$

$$
= 0
$$

Hence welfare is decreasing in $q$. \qed

Claim 3. Default probability is increasing in $q$.

Proof. Note $\mathbb{E} [Pr (Default)] = \mathbb{E} [D^{agg}]$ and the claim follows directly from claim 1. \qed

240
Claim 3. $\mathbb{E}[D^{agg}] / \mathbb{E}[I^{agg}]$ is increasing in $q$.

Proof. Follows directly from claims 1 and 2. \hfill \Box

Claim 4. $\lim_{q \to 1} V(0) = 0$.

Proof. We can write the ex-ante welfare as

$$V(0) = \frac{1 + 2\alpha (q - 1) - 2q + z}{4q}$$

Using the fact that $\lim_{q \to 1} z = 1$ and taking limit, the result is immediate. \hfill \Box

Proof of Proposition 3.5

Take $\alpha_1 > \alpha_2$. From lemma 3.2, we have $\ell(b, q, \alpha_1) = \ell(b, q, \alpha_2)$ and that $b(\ell, q, \alpha)$ is increasing in $\alpha$. This means that the downward sloping borrower best response curve in Figure 3-3 shifts upwards as $\alpha$ increases, as the figure illustrates. Since the lender's best response slopes upwards ($\ell_b(b, q, \alpha) > 0$, also from Lemma 3.2), it follows that $b^*(q, \alpha_1) > b^*(q, \alpha_2)$ and $\ell^*(q, \alpha_1) > \ell^*(q, \alpha_2)$.

Next, from lemma (3.4) we have

$$\mathbb{E}[D^{agg}] = \frac{1 - b^{-1}}{1 - q b^{-1}}$$

$$\mathbb{E}[D^{agg}] / \mathbb{E}[I^{agg}] = \frac{1 - b^{-2}q}{b^{-1}(1 - q)}$$

both of which are increasing in $b$, which establishes the result that they are also both increasing in $\alpha$. \hfill \Box

Proof of Proposition 3.6

From lemma (3.5) we have

$$\mathbb{E}[I^{agg}] = \frac{(\alpha - 1)(z + 1 - 2\alpha(1 - q))}{2\alpha q (2\alpha - 1)}$$
Differentiating with respect to $\alpha$, we get

$$
\frac{dE[I_{agg}]}{d\alpha} = \frac{q(2\alpha - 2\alpha^2z - 10\alpha^2 + 8\alpha^3) - (z - 6\alpha + 4\alpha^2z + 12\alpha^2 - 8\alpha^3 - 4\alpha z + 1)}{2\alpha^2q(2\alpha - 1)^2z}
$$

From this expression, one can verify that for any given $q \in (0, 1)$, there exists an unique $\alpha(q) \in (1, \infty)$ such that

$$
\frac{dE[I_{agg}]}{d\alpha} \begin{cases} < 0 & \text{for } 1 \leq \alpha < \alpha(q) \\ > 0 & \text{for } \alpha > \alpha(q) \end{cases}
$$

and

$$
\frac{d\alpha(q)}{dq} < 0.
$$

Proof of Proposition 3.7

Consider projects 1 and 2 as represented in figure 3-6. Project 2 would yield the borrower a value of $V(0; \alpha^H, q)$, whereas project 1 would yield him

$$
U^{SL}(\alpha^L) = \alpha L^{SL}(\alpha^L) + \mathbb{E}\left[\min\left(\tilde{c}, D^{SL}\right)\right]
$$

By construction, $\alpha^H$ were chosen such that $V(0; \alpha^H, q) = USL(\alpha^L)$. Since $U^{SL}(\alpha)$ is increasing in $\alpha$, any linear-flat project of the form

$$
R(I) = \alpha \min\left(I, L^{SL}(\alpha)\right)
$$

where $\alpha^H > \alpha > \alpha^L$ would yield strictly higher utility to the borrower than project 2, despite the fact that for any given level of investment, the linear-flat project yields lower output than the project 2:

$$
\alpha^H I > \alpha \min\left(I, L^{SL}(\alpha)\right) \quad \text{for all } I > 0
$$

Proof of Proposition 3.8

Proof of Claim 1. For an increasing and concave investment function $R(\cdot)$ with $R'(0) > 1$, the first-order condition in equation (3.3.1) that characterizes the single-lender equilibrium can be
re-written as
\[ R'(p(D^*) D^*) \times [p(D^*) + p'(D^*) D^*] = p(D^*) \]

If an interest rate cap were set to be \( 1 + \tilde{r}^{SL} = \frac{1}{p(D^*)} \), the borrower could propose to pledge \( D^* \) and raise \( p(D^*) D^* \) from the very first lender. No future lender would be willing to provide additional investment to the borrower because doing so would require an interest rate higher than \( 1 + \tilde{r}^{SL} \) to break even, but such a rate is prohibited by the interest rate cap. Hence the full commitment allocation can be achieved under \( \tilde{r}^{SL} \).

When \( R(I) = \alpha I \), the first-order condition simplifies to
\[ \alpha (1 - 2D^*) = 1 - D^* \]
hence \( D^* = \frac{\alpha - 1}{2\alpha - 1} \). The optimal interest cap is thus
\[ 1 + \tilde{r}^{SL} = \frac{1}{1 - D^*} = 1 - \frac{1}{\alpha} \]

Proof of Claim 2. Consider an interest rate cap of \( \tilde{r} \). As explained in Claim 1, the maximum debt that can be pledged is

Proof of Claim 3. When \( \tilde{r} < \tilde{r}^{SL} \), the interest rate cap is inefficiently low and the borrower can pledge less debt facevalue than he would have done under full commitment. The unique equilibrium under the interest rate cap would involve the borrower pledging \( D = 1 - \frac{1}{1 + \tilde{r}} \) debt and raising \( (1 - D) D \) investment from the very first lender. In the extreme case where \( \tilde{r} = 0 \), the borrower would be unable to raise any investment from the lenders, achieving an even lower level of welfare than under the unregulated equilibrium.

Proof of Claim 4. When \( \tilde{r} > \tilde{r}^{SL} \), the full commitment allocation is unattainable as with probability \( q \) the borrower will meet the second lender and pledge a strictly positive amount of debt for any level of outstanding debt below one.

Next we show that for \( \tilde{r} < \infty \) the cap unambiguously improves expected investment and welfare while lowering expected debt and interest rate relative to the unregulated equilibrium. Using the techniques created in the proof for proposition 3.3, for any game with finite lenders \( K \) we can find a sequence of aggregate debt \( \{D^K_1, ..., D^K_K\} \) where \( D^K_i \) corresponds to the aggregate debt level had the borrower reach lender \( i \) in a game with total lender \( K \), where lender indices start backwards with the last lender being lender 1. Using a simple perturbation argument, we know that for \( K \) such that \( D^K_1 < \bar{D} \leq D^K_{K+1} \), the infinite lender game with debt cap \( \bar{D} \) would have a unique SPE where the borrower reaches the debt cap when borrowing from \( (K + 1) \)-th lender. Furthermore, using the same recursive definition of lender and borrower strategies in equilibrium as we adopted
in proposition 3.3, it is clear that borrower's ex-ante expected investment, aggregate debt level, and welfare with debt cap $\bar{D}$ is in between the corresponding equilibrium quantities for the finite lender games with $K$ and $K + 1$ lenders.
Bibliography


