INTERMEDIATION VARIETY*

Jason Roderick Donaldson† Giorgia Piacentino‡ Anjan Thakor§

May 11, 2020

Abstract

We explain why banks and non-bank intermediaries coexist in a model based only on differences in their funding costs. Banks enjoy a low cost of capital due to safety nets and money-like liabilities. We show that this can actually be a disadvantage: it generates a soft-budget-constraint problem that makes it difficult for banks to credibly threaten to withhold additional funding to failed projects. Non-banks emerge to solve this problem. Their high cost of capital is an advantage: it allows them to commit to terminate funding. Still, non-banks never take over the entire market, but coexist with banks in equilibrium.

*We gratefully acknowledge helpful comments from Philip Bond (the Editor), an anonymous Associate Editor, two anonymous referees, Mike Burkart, Francesca Cornelli, Giovanni Dell'Ariccia, Philip Dybvig, Daniel Ferreira, Radha Gopalan, Vincent Glode, Denis Gromb, Ohad Kadan, Nataliya Klimenko, Mina Lee, Adrien Matray, Asaf Manela, Alan Morrison, Balazs Szentes, Basil Williams, Andy Winton, Vijay Yerramilli, Kostas Zachariadis, Jean-Pierre Zigrand and seminar participants at the 2015 EFA Summer Meetings, the 2015 FIRS Conference, the 2017 Five Star Conference, the London School of Economics, the 2018 SFS Cavalcade, and Washington University in St. Louis. Giuditta Perinelli and Xiaobo Yu provided outstanding research assistance. We alone are responsible for any remaining errors.

†Washington University in St. Louis and CEPR; jasonrdonaldson@gmail.com

‡Columbia University, CEPR, and NBER; g.piacentino@gsb.columbia.edu

§Washington University in St. Louis and ECGI; thakor@wustl.edu

Depository financial institutions—"banks"—have a low cost of capital,[1] most likely because their liabilities benefit from a moneyness premium and government safety nets. Perhaps due to this funding-cost advantage over non-banks, banks provided the bulk of finance until the late 1970s, when deregulation removed barriers to entry and incumbent banks faced increased competition from new entrants.[2] At this point, non-depository financial institutions—"non-banks"—proliferated. As Remolona and Wulfekuhler (1992) put it,

> During the 1980s, U.S. commercial banks faced increased competition in their lending activities from other financial intermediaries...[which] enjoyed their success despite carrying apparently heavier capital burdens and lacking the advantage of deposit insurance (p. 25).

It seemed like non-banks, such as venture capitalists and finance companies, could even replace banks. Indeed, in 1994, Boyd and Gertler wrote a paper called "Are Banks Dead?" which begins

> It is widely believed that in the United States, commercial banking is a declining industry [because] nonbank credit alternatives have grown rapidly over the last 15 years (p. 2).

But banks remain alive and well today, with over $12.5 trillion in deposits in the US.[3] Nonetheless, non-banks do compete with banks, overcoming their funding cost disadvantage and providing a substitute form of business finance.

Non-banks differ from banks in a number of ways. Compared to banks, they (i) finance different kinds of entrepreneurs; notably, they are particularly likely to finance start-ups and other innovative entrepreneurs, which are associated with relatively high agency costs due to imperfect information or misaligned incentives. Further, they (ii) charge entrepreneurs relatively high rates, (iii) have relatively short-term relationships, (iv) are relatively intolerant of failure, (v) exist in relatively competitive financial markets, and (vi) are relatively scarce. Non-bank-financed entrepreneurs also differ from bank-financed ones. Compared to bank-financed entrepreneurs, those financed by non-banks (vii) do relatively high-agency-cost projects and (viii) are relatively unlikely to obtain financing from other financiers (see Section 4.2 for references). What explains this variety in bank vs. non-bank finance?

We develop a model that suggests that all of these differences between bank and non-bank finance could result from a single source: heterogeneity in financiers' cost of

[1]For example, Startz (1979) and Nagel (2016, online Appendix B) estimate that deposit rates are one-third to one-half of the competitive rate. See, e.g., Diamond (2019), Donaldson and Piacentino (2019), and Merton and Thakor (2018) for theories of moneyness premiums.

[2]See, e.g., Stiroh and Strahan (2003) on how late-1970s deregulation removed "restrictions...shielding banks from outside competition...[and] created a more competitive environment" (p. 801).

[3]See the FDIC Quarterly Banking Profile at https://www.fdic.gov/bank/statistical/stats/2018dec/industry.pdf.

capital.

Model preview. In the model, financiers enter and become either banks or non-banks. Banks and non-banks are identical in every way but one: banks have a lower cost of capital. Each financier meets an entrepreneur, who chooses one of two types of project to seek financing for. Both types can last for up to two stages—if a project fails at a first stage, it is either terminated or refinanced for a second stage—and both require (unobservable) effort at each stage to be profitable. The difference between the two types is that the cost of effort can be high or low, hence we call the projects "high agency cost" (HAC) or "low agency cost" (LAC).

To incentivize effort, financiers must either leave the entrepreneurs "agency rents" in the event of success or punish them with termination in the event of failure. It is cheaper for financiers to rely on the termination threat. However, the termination threat might not be credible due to the soft-budget-constraint problem inherent in staged financing.[4] That is, even if financiers would like to commit ex ante to terminate, ex post they might prefer not to—colloquially, you might want to throw good money after bad, even if you would have liked to commit not to. We assume that absent a credible termination threat, only LAC projects are viable, because HAC projects are too costly to finance net of agency rents, even though they could have higher total value.

Result preview. Our first main result is that non-banks' high cost of capital can be an advantage. The reason is that it makes refinancing failed entrepreneurs unattractive, and thus creates a credible termination threat. In other words, non-banks, with their high cost of capital, can harden soft budget constraints, making financing HAC projects profitable. In contrast, banks, with their low cost of capital, cannot, making financing HAC projects inviable. This indirect advantage of non-banks' high-cost of capital counteracts their direct funding-cost disadvantage, and allows them to compete with banks. Indeed, we find that non-banks co-exist with banks and that only non-banks fund HAC projects in equilibrium (fact (i)). Moreover, because the termination threat reduces the agency rent they must surrender to entrepreneurs, they can charge higher rates (fact (ii)). But, because, unlike banks, they terminate, rather than refinance, failed projects, their relationships with entrepreneurs are relatively short term (facts (iii) and (iv)).

This result echoes Bolton and Scharfstein (1990) in that "committing to terminate funding if a firm's performance is poor...mitigate[s] managerial incentive problems" (p. 93). Our insight is that it is a financier's own high cost of capital that makes this commitment credible. Thus, as in, e.g., Jensen and Meckling (1976) and Zwiebel (1995), debt disciplines entrepreneurs, but, unlike in these papers, it is not debt on entrepreneur's own balance sheets, but rather that on their financiers' balance sheets.

[4] See, e.g., Dewatripont and Maskin (1995) and Kornai (1979, 1980).

The result also resonates with practice. Non-banks' ability to harden soft budget constraints is arguably their main disciplining tool. As Sahlman (1990) stresses for VCs, "the credible threat to abandon a venture, even when the firm is economically viable, is the key to the relationship between the entrepreneur and the VC" (p. 507). And banks' inability to harden soft budget constraints was a first-order concern for economists worried about the decline of banking. As Jensen (1989) puts it, "banks' chief disciplinary tool, their power to withhold capital from...companies, has been vastly reduced."

Our second main result is that non-banks' hard budget constraints decrease not only the agency rents they need to leave entrepreneurs, but also the competition they face from other financiers. In fact, rival financiers are unwilling to provide refinancing to non-bank financed entrepreneurs and entrepreneurs are captive to incumbent non-banks as a result (fact (viii)). The reason is that imposing a hard budget constraint allows a non-bank to demand high repayments and a rival will not finance an entrepreneur who already owes a high repayment to someone else.

This is the counterpart to results in the literature on how competition among financiers can affect entrepreneurs' incentive problems (see Boyd and De Nicolò (2005) and Martinez-Miera and Repullo (2010)). Our insight is that, if, unlike in these papers, competition among financiers is non-exclusive—different financiers can fund the same entrepreneur at the same time—then the mechanism can also work in the other direction. Specifically, mitigating agency problems (via the termination threat) can effectively decrease competition among financiers because it allows them to charge entrepreneurs such high rates initially that no one else wants to fund them subsequently.

Our third main result is that entrepreneurs choose their projects based on the kind of finance they have access to. Because they enjoy agency rents, entrepreneurs prefer HAC projects. But they do not want to choose projects that they will be unable to finance. Thus, knowing that only non-banks, with their use of hard budget constraints, will finance HAC projects, entrepreneurs choose HAC projects when they have access to non-bank finance and LAC projects when they do not (fact (vii)).

Like some other papers in the literature (discussed below), there is sorting between financiers and projects—in equilibrium, banks and non-banks finance different types of projects. Unlike in this literature, however, entrepreneurs are ex ante identical. Thus, the mix of financiers in the market determines the mix of projects, not the other way around.

Our fourth main result is that non-banks become more important as competition among financiers increases. Non-banks enter only competitive markets and provide an increasing proportion of financing as competition increases. However, they do not take over the whole market, possibly remaining scarce for all levels of competition (facts (v) and (vi)). To understand the result, first observe that competition does not affect non-banks, which can keep entrepreneurs captive—they are always effective monopolists—but does affect banks, which cannot. If competition is low, this benefit of monopoly power for non-banks is not enough to outweigh the direct disadvantage of their high cost of capital, and all financiers specialize in traditional banking. As competition increases, non-banking becomes attractive, and some financiers specialize in it to exploit the high rates they can charge to captive entrepreneurs. Not all financiers do this, however. Some always specialize in banking. The reason is that if everyone were to specialize in non-banking, then there would be a paucity of banks and we would be back in the case of low competition among banks, so banking would become attractive again. Hence, banks co-exist with non-banks, even for very high competition.

In summary, for low competition, there are only banks, and entrepreneurs choose LAC projects. For higher competition, banks finance only LAC projects, and non-banks emerge to fund HAC projects, providing an increasing share of finance as competition increases, but possibly remaining scarce, even in the perfect competition limit. (See Figure 1.3.)

This connection between competition and the mix of financiers in the market is new to the literature. It arises through an externality that one bank's entry imposes on other banks, but not on non-banks: by increasing competition for continuation financing, bank entry makes soft-budget-constraint problems worse. This harms banks, but not non-banks, because only banks suffer from soft-budget-constraint problems.

Figure 1: Financing Regimes as a Function of Competition among Financiers

For low competition	For higher competition	As competition increases
All financiers specialize in banking	Most financiers specialize in banking	Some financiers continue to specialize in banking
None specialize in non-banking	But a few specialize in non-banking	More and more, but never all, specialize in non-banking
All entrepreneurs are bank-financed and choose LAC projects	Bank-financed entrepreneurs choose LAC projects	Bank-financed entrepreneurs still choose LAC projects
There are no non-bank-financed entrepreneurs	Non-bank-financed entrepreneurs choose HAC projects	Non-banked financed entrepreneurs still choose HAC projects

Further results. We explore three extensions. In the first, we relax the assumption that financing HAC projects is prohibitively costly for banks. Thus, entrepreneurs could choose HAC projects, regardless of the type of finance they have access to. We show, however, that they may still choose the LAC project when they have access to banks. Specifically, for high competition, the observed behavior of entrepreneurs is qualitatively unchanged from our baseline results: entrepreneurs who meet banks choose LAC projects and entrepreneurs who meet non-banks choose HAC project. In the second extension, we suppose that there is congestion among similar financiers, e.g., because they look for similar entrepreneurs. In this case, the more non-banks enter, the harder it is for other non-banks to find entrepreneurs to finance, and likewise for banks. Whether this makes non-banking or banking more attractive depends on which financiers are most affected by the congestion; hence it can either increase or decrease the proportion of non-banks that operate in equilibrium. Either way, however, it does not qualitatively change our results. In the third extension, we suppose that there is a limited supply of HAC projects. In this case, the more non-banks enter and fund HAC projects, the fewer HAC projects are left for other non-banks to fund. This makes non-banking less attractive; hence it decreases the proportion of non-banks that operate in equilibrium. However, it does not qualitatively change our results.

Related literature. Our paper contributes to the literature on how borrowers choose between competing sources of finance, most of which focuses on the the trade-off between bank and market finance.[5] In this literature, borrowers are typically endowed with heterogenous projects which determine whether it is advantageous for them to seek bank or market finance. Banks typically have an informational advantage over markets by assumption, as in Diamond (1991), Holmström and Tirole (1997), and Rajan (1992), for example. Information-sensitive borrowers thus choose banks to benefit from bank monitoring or flexibility, whereas borrowers less in need of monitoring choose markets to avoid compensating banks for monitoring or giving them information rents. In Boot and Thakor (1997), the trade-off is between the market's ability to aggregate information and banks' ability to resolve moral hazard. Thus, again, borrowers' exogenous characteristics determine their choice of financing source. Unlike this literature, we assume that borrowers are ex ante identical. Differences among them arise ex post based on their source of finance. And, also unlike this literature, we focus on the trade-

[5]See Allen and Gale (2004), Besanko and Kanatas (1993), Bolton and Freixas (2000), Chemmanur and Fulghieri (1994), Gersbach and Uhlig (2007), Hoshi, Kashyap, and Scharfstein (1993), Rajan (1992), Repullo and Suarez (2000), Song and Thakor (2010), and von Thadden (1999). Many papers study competition among banks, rather than between banks and other sources of finance; see, e.g., Boot and Thakor (2000), Boyd and De Nicolò (2005), Cao and Shi (2000), Cetorelli (2004), Dell'Ariccia (2000), Dell'Ariccia and Marquez (2004), Guzman (2000), Hellmann, Murdock, and Stiglitz (2000), Keeley (1990), Marquez (2002), Martinez-Miera and Repullo (2017), Matutes and Vives (2000), Petersen and Rajan (1995), Repullo (2004), Sharpe (1990), and Wagner (2009).

off between bank and non-bank finance, rather than between bank and market finance. This is likely to be the most relevant trade-off for the kinds of innovative/entrepreneurial borrowers we model.

There are a few other papers in which banks coexist with other types of financiers. In Bond (2004), they coexist with less-diversified financiers, such as conglomerates, which can economize on the costs of information sharing in some circumstances. In Ueda (2004), they coexist with VCs, which can screen entrepreneurs' projects better, but cannot commit not to expropriate them.[6] In Begenau and Landvoigt (2017) and Chrétien and Lyonnet (2019), they coexist with shadow banks, which are less regulated, but do not benefit from cheap funding due to moneyness or deposit insurance. In Hanson, Shleifer, Stein, and Vishny (2015), they also coexist with shadow banks, which, in line with the other papers cited, are less regulated. However, in contrast with the other papers, these shadow banks also enjoy a low cost of capital from creating money-like liabilities. Thus, they do not resemble the non-banks in our model, but are closer to our banks, whose defining feature is their low cost of capital. Indeed, their non-banks are closest to money market mutual funds, which invest in marketable securities, whereas ours are closest to venture capitalists or finance companies, which finance early-stage entrepreneurs.

Our model is also related to models of the market for venture capital, which also stress staged financing. Like us, Inderst and Mueller (2004), Jovanovic and Szentes (2013), Khanna and Mathews (2017), and Nanda and Rhodes-Kropf (2013) use models of bilateral meetings to embed dynamic VC-entrepreneur relationships in a wider market.[7] Many of these papers include search-and-matching frictions, which are likely to be first order for early-stage entrepreneurs with hard-to-assess projects.[8] We can too, but we do not have to for our results. What matters is that we can capture scarcity, not search frictions. Namely, some entrepreneurs can go unfunded just because capital is scarce, even if matching is frictionless. This is certainly first order for potential entrepreneurs, who report that raising capital is their principal problem (Blanchflower and Oswald (1998)).

Layout. Section 1 presents the model. Section 2 analyzes the bilateral contracting problem between entrepreneurs and financiers. Section 3 solves for the equilibrium and presents our main result on intermediation variety. Section 4 discusses our assumptions, empirical content (including evidence for the facts listed at the beginning), and policy implications. Section 5 develops extensions. Section 6 concludes. All proofs are in the

[6]See also Chan, Siegel, and Thakor (1990). In that paper, banks, VCs, and markets all coexist, and a borrower's financing choice depends on his experience and reputation.

[7]Some other papers, e.g., Boualam (2018), Payne (2018), Donaldson, Piacentino, and Thakor (2019), Herkenhoff (2019), and Wasmer and Weil (2004), use related models to study the market for bank credit.

[8]Indeed, in one survey, 20% of aspiring entrepreneurs say that *where* to get finance is their *biggest* concern (Blanchflower and Oswald (1998)).

Appendix.

1 Model

Time is discrete and the horizon is infinite. Overlapping generations of entrepreneurs seek financiers to provide capital to two-stage projects. The projects suffer from a soft-budget-constraint problem, requiring additional capital if they do not pay off at the initial stage. Incumbent financiers have a monitoring advantage, but still compete with the next generation of financiers to provide continuation capital. This competition is the only link between generations. (We omit time indices since we focus on stationary equilibria.)

1.1 Entrepreneurs and Projects

At each date, a unit continuum of identical, penniless, risk-neutral entrepreneurs is born. Each entrepreneur meets a financier with probability Q, which reflects the supply of financiers relative to entrepreneurs. We take Q as our measure of competition among financiers. If an entrepreneur meets a financier, he may raise capital to invest in one of two projects. Each is associated with agency problems, but one has higher agency costs than the other. Hence, we index the projects by the level of agency costs, with $\alpha = A$ denoting the high-agency-cost (HAC) project and $\alpha = a$ denoting the low-agency-cost (LAC) one. If an entrepreneur does not meet a financier, he gets a reservation payoff normalized to zero.

The projects resemble those in Crémer (1995). Each project lasts two stages with moral hazard at each stage and a soft budget constraint. Specifically, each project requires first-stage financing K_0 at the initial date and continuation financing K_1^α at the interim date if it does not succeed at the first stage. If the project succeeds (at either stage), it pays off y^α. Otherwise, it pays off nothing. The probability of success, denoted by π_1 at the first stage and π_2 at the second, depends on the entrepreneur's effort. If he works, the project succeeds with probability p; if he shirks, it succeeds only with probability $p - \Delta$. Although working increases the expected payoff of the project, it is costly for the entrepreneur, because it entails forgoing (non-pecuniary) private benefits β^α at each stage, where $\beta^\alpha = B^\alpha$ unless the financier monitors the project, in which case it is reduced to $\beta^\alpha = b^\alpha$ in the second stage, as discussed below.

A project $\alpha \in \{a, A\}$ is thus characterized by seven parameters K_0, K_1^α, y^α, p, Δ, B^α, and b^α. Observe, however, that only the payoff, second-stage financing cost, and private benefits depend on its type α (although below we often omit the superscript α even from these parameters). Limiting the parameters that vary across the projects imposes discipline (limiting our free parameters) and also simplifies the equations. It also means that we abstract from some important ways in which projects can differ, e.g., in their riskiness. But, as we discuss in Section 4.1, the model can easily be adapted to capture such differences.

Later, we make parametric assumptions, which define how LAC and HAC projects differ (Section 2.3). For now, we just assume that the payoff y is sufficiently large for both types.

Assumption 1. *Projects' payoffs are sufficiently high:*

$$y \geq \frac{1}{\Delta^2} \max\left\{\frac{p(1+p)B - \Delta pb - \Delta^2 K_1}{(1-p)}, pB\right\}. \tag{1}$$

As we show in Appendix B, this assumption ensures it is always better for financiers to offer repayments so that the entrepreneur works at both stages. Roughly speaking, working makes it more likely to get y; hence, working is optimal as long as y is large enough (relative to the agency and financing costs captured by the RHS of inequality (1)).

1.2 Financiers

At each date, a continuum of identical risk-neutral financiers is born. Each chooses to become either a bank or a non-bank and meets an entrepreneur with probability q, which is a decreasing function of the number of entrepreneurs, $q'(Q) < 0$. We let φ denote the (endogenous) proportion of non-banks.

The only difference between financiers is their cost of capital ρ. Banks have a low cost of capital, which we normalize to zero ($\rho = 0$), relative to non-banks, which have a higher cost of capital ($\rho = r > 0$). This cost of capital defines the hurdle rate that financiers use to discount their own investments.

Both types of financiers want to invest in entrepreneurs' projects. If a financier does not meet an entrepreneur, it exits, getting a reservation payoff normalized to zero. If a financier meets an entrepreneur, it can make the entrepreneur a take-it-or-leave-it offer of initial financing K_0 in exchange for the repayment R_1 in the event that the project succeeds at the first stage. If the entrepreneur does not succeed at the first stage, the financier can make a take-it-or-leave-it offer of continuation financing K_1 in exchange for the additional repayment R_2 in the event that the project succeeds in the second stage.[9] If the entrepreneur rejects this offer, he can try to find continuation financing from a rival financier in a market populated by the next generation of financiers. We

[9]We assume that R_1 is prioritized ahead of R_2. However, this does not matter for the qualitative results. The reason is that, as long as it is optimal to incentivize the entrepreneur to work at both stages (cf. Assumption 1), repayments are determined by the entrepreneur's IC, which depends on only the total stock of debt, hence not on the relative priority of the debts that make it up. See equation (7) below.

let $\hat{Q}$ denote the probability the entrepreneur meets a rival that offers continuation financing and $\hat{R}_2$ denote the repayment the rival offers. If the entrepreneur does not get financing, the project is scrapped, paying off zero.

We assume that incumbent financiers have an advantage in monitoring entrepreneurs, due to, say, proprietary information they acquire about the entrepreneur, as in Rajan (1992).[10] Following Holmström and Tirole (1997), we assume that if an entrepreneur gets continuation financing from his incumbent financier, his second-stage private benefits are reduced from B to b, but if he gets it from a rival his private benefits are still B.[11]

1.3 Timeline

At each date, for a given level of competition among financiers Q, each financier chooses to be a bank or a non-bank. It then meets an entrepreneur with probability q, which is a decreasing function of Q. Symmetrically, each entrepreneur meets a financier with a probability that is proportional to the number of financiers of that type (banks and non-banks), meeting a bank with probability $(1-\varphi)Q$ and a non-bank with probability φQ. After meeting a financier, the entrepreneur chooses a project to seek financing for. Then, the financier offers the entrepreneur financing terms. After this, the entrepreneur works or shirks, and the project either succeeds or fails. If it succeeds, the entrepreneur makes the agreed repayment. If it fails, the entrepreneur does not make the repayment and the sequence repeats: the financier makes an offer to fund the continuation of the project, the entrepreneur works or shirks, and the project succeeds or fails. If it succeeds, the entrepreneur makes the agreed repayment; otherwise, he repays nothing. Entrepreneurs and financiers exit if they do not meet anyone.[12]

Importantly, a financier which offers continuation financing takes into account the fact that an entrepreneur rejecting it can try to find continuation financing from a rival financier with probability $\hat{Q}$. Recall that the entrepreneur's private benefits are lower

[10] Botsch and Vanasco (2019) find empirical evidence of banks' "learning by lending," by which incumbent financiers obtain such an informational advantage over competitors, and Nakamura and Roszbach (2018) find evidence of their monitoring.

[11] Formally, this can be modeled by assuming that incumbent financiers have a lower cost of monitoring than rivals. Indeed, it is equivalent to the extreme assumption in Crémer (1995), that the monitoring cost is zero if the incumbent financier finances the second stage, but infinite if a new financier does. Realistically, both incumbents and rivals are likely to monitor entrepreneurs. This is just a simple way to capture the idea that incumbents can monitor more cheaply, and hence are likely to monitor more. (See Section 4.1 for a discussion.)

[12] This assumption that everyone gets only one chance to match keeps the model stationary, so that competition is the same at each period. We intentionally abstract from dynamics, using the OLG set-up just to capture the effect of competition on multi-stage financing in a simple way. See Biais and Landier (2015) for a model in which a similar link between overlapping generations of entrepreneurs does matter for aggregate dynamics.

with the incumbent financier, given its monitoring advantage.

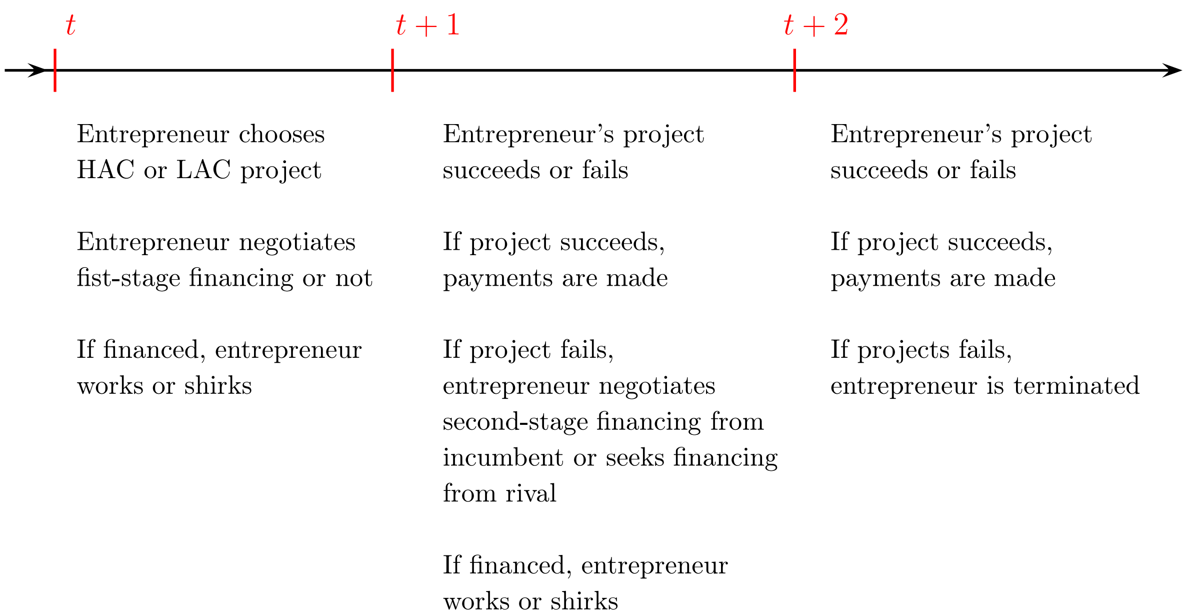

Figure 2: Timeline given a match between a financier and an entrepreneur

2 Contracting Problem

In this section, we analyze the two-stage bilateral contracting problem between a financier and an entrepreneur. First, we set up the contracting problem in terms of participation constraints (PCs) and incentive constraints (ICs).

Second, we combine the PCs and the ICs above to derive four results: (i) a condition for the entrepreneur to face a soft budget constraint (SBC), (ii) a condition for him to be captive to his incumbent financier, (iii) a condition under which an entrepreneur who faces a hard budget constraint (HBC) is captive, and (iv) an expression for his continuation payoff (i.e. his payoff given failure at the first stage).

Finally, in light of these results, we impose parametric assumptions that ensure that the types of projects/financiers are sufficiently different and that whether an entrepreneur faces an SBC and/or is captive depends on his project/financier.

2.1 Participation and Incentive Constraints

Here we analyze the two-stage contracting problem implied by the set-up above, writing first a financier's PCs to provide finance at each stage and then an entrepreneur's ICs to work at each stage.

2.1.1 Financiers' participation constraints

We start with the second-stage PCs. The entrepreneur arrives at the second stage only if his project fails at the first. The incumbent financier can offer him continuation financing in exchange for an additional repayment, but the entrepreneur can reject it and look for a rival. Financiers offer contracts only if they satisfy their PCs. The PCs are different for the incumbent financier and the rival, because the incumbent takes into account that (i) providing continuation financing makes it more likely that it will recoup the initial repayment and (ii) incentivizing effort is relatively cheap given its ability to monitor. We describe these PCs in turn.

The incumbent financier is willing to provide continuation financing if its cost K_1 is lower than the discounted expected value of its total repayment $R_1 + R_2$:

$$K_1 \leq \pi_2 \frac{R_1 + R_2}{1 + \rho}. \tag{2}$$

If this is satisfied for some feasible repayment $R_1 + R_2$, equilibrium success probability π_2, and the incumbent's cost of capital ρ, then we say there is a *soft budget constraint*, which we denote by $\mathbb{1}_{\{\text{SBC}\}} = 1$; otherwise, we say there is a *hard budget constraint*, which we denote by $\mathbb{1}_{\{\text{SBC}\}} = 0$.

A rival financier is willing to provide continuation financing if its cost K_1 is lower than the discounted expected value of its repayment $\hat{R}_2$:

$$K_1 \leq \pi_2 \frac{\hat{R}_2}{1 + \rho}. \tag{3}$$

If this is satisfied for some feasible repayment $\hat{R}_2$, equilibrium π_2, and a rival's cost of capital ρ, then we say the entrepreneur is *not captive*, so the probability of getting continuation finance from a rival is positive, $\hat{Q} > 0$; otherwise, we say that he is *captive*, so $\hat{Q} = 0$.

We now turn to the first-stage PC. We streamline the exposition here by restricting attention to the case in which the entrepreneur gets continuation financing from his incumbent financier if he gets it at all, which turns out to be the only relevant case (see Corollary 1 in the Appendix). [13] However, the possibility of getting continuation financing from a rival financier can be a relevant outside option for the entrepreneur, and it affects the terms that the incumbent offers.

At the first stage, a financier takes into account that it could refinance the en-

[13] Basically, if a rival financier, which gets only $\hat{R}_2$ and cannot monitor, is willing to provide finance, so is the incumbent, which gets $R_1 + R_2$ and can monitor .

trepreneur at the second stage, i.e. that it could have a soft budget constraint. The financier's PC reads:

$$\pi_1 \frac{R_1}{1+\rho} + (1-\pi_1)\mathbb{1}_{\{\text{SBC}\}}\, \pi_2 \frac{R_1+R_2}{(1+\rho)^2} \geq K_0 + (1-\pi_1)\mathbb{1}_{\{\text{SBC}\}} \frac{K_1}{1+\rho}, \tag{4}$$

where, given the financier's cost of capital ρ, the LHS is the present value of the entrepreneur's repayments and the RHS is the present value of the financier's capital outlay.

2.1.2 Entrepreneurs' incentive constraints

We start with the second-stage IC. An entrepreneur who owes his financier(s) $R_1 + R_2$ prefers to work than to shirk if

$$p\big(y - R_1 - R_2\big) \geq (p - \Delta)\big(y - R_1 - R_2\big) + \beta. \tag{5}$$

The LHS is his expected payoff if he works—his success probability is $\pi_2 = p$. The RHS is his expected payoff if he shirks—his success probability is only $\pi_2 = p - \Delta$—but he gets private benefits β, where $\beta = b$ if he gets continuation financing from his incumbent and hence is monitored, and $\beta = B$ if he gets it from a rival. This IC can be rewritten as an upper bound on his total repayment $R_1 + R_2$:

$$R_1 + R_2 \leq y - \frac{\beta}{\Delta}. \tag{6}$$

We now turn to the first-stage IC, which depends on his continuation value given failure, denoted by u_1. Specifically, the entrepreneur prefers to work than to shirk if and only if

$$p\big(y - R_1\big) + (1-p)u_1 \geq (p - \Delta)\big(y - R_1\big) + (1 - p + \Delta)u_1 + B. \tag{7}$$

The LHS is his expected payoff if he works—his success probability is $\pi_1 = p$. The RHS is his expected payoff if he shirks—his success probability is only $\pi_1 = p - \Delta$—but he gets private benefits $\beta = B$. This IC can be rewritten as an upper bound on his repayment R_1:

$$R_1 \leq y - \frac{B}{\Delta} - u_1. \tag{8}$$

Note that increasing the entrepreneur's continuation value u_1 decreases the repayment the financier can extract, because the financier must leave him more rent to incentivize him to work. u_1, in turn, depends on whether the entrepreneur can fund continuation at the second stage, and at what terms.

2.2 Soft Budget Constraint (SBC), Captivity, and Continuation Value

We now use the preceding analysis to derive results on the outcome at the interim date.

Combining the incumbent financier's second-stage PC and the entrepreneur's second-stage IC gives a condition for the entrepreneur to face a soft budget constraint:

Proposition 1. (Soft budget constraint.) *An incumbent financier provides continuation financing if and only if the following condition holds:*

$$p\left(y - \frac{b}{\Delta}\right) \geq (1+\rho)K_1, \tag{9}$$

i.e. if and only if its cost of capital ρ is sufficiently low.

The condition of the lemma (inequality (9)) says that the maximum that incumbent financiers can extract at the second stage—here, the total expected payoff py minus the agency rent pb/Δ—needs to exceed the cost of continuation compounded for one period at the incumbent's cost of capital ρ. Hence, increasing ρ makes it harder to satisfy inequality (9). Intuitively, because a financier's cost of capital defines the hurdle rate it applies to its own investments, increasing it enough leads the financier to deny financing to entrepreneurs at the continuation stage.

Combining the incumbent financier's second-stage PC and the entrepreneur's second-stage IC for a given the first-stage repayment, R_1, yields a condition for the entrepreneur to be captive to his incumbent financier:

Proposition 2. (Endogenous captivity.) *An entrepreneur cannot obtain continuation financing from a competing financier—i.e. he is captive to his incumbent financier—if and only if the following condition holds:*

$$(1+\rho)K_1 > \max\left\{p\left(y - \frac{B}{\Delta} - R_1\right), (p-\Delta)(y-R_1)\right\}, \tag{10}$$

for $\rho \in \{0, r\}$, i.e. if and only if his initial repayment R_1 is sufficiently high.

The condition in the proposition (inequality (10)) says that a rival financier is unwilling to provide continuation financing if the maximum that it can extract from the entrepreneur (the RHS of inequality (10)) is lower than the cost of continuation, which is its capital outlay compounded for one period at the rival's cost of capital ρ (the LHS of inequality (10)). The next corollary follows from combining the results above (Proposition 1 and Proposition 2):

Corollary 1. *Suppose*

$$K_1 > (p-\Delta)\frac{B}{\Delta}. \tag{11}$$

Then, if the entrepreneur has a hard budget constraint, he is captive.

The results so far suggest that a high cost of capital, which seems like a disadvantage, could actually be an advantage:[14] it allows the financier to impose a hard budget constraint (Proposition 1), and an entrepreneur who faces a hard budget constraint is endogenously captive to his incumbent. To see why, there are two steps. The first is to observe that if the incumbent can require a high R_1, then he can keep the entrepreneur captive (Proposition 2). The reason is that the more he owes to his incumbent from the first stage, the less he can promise to a rival at the second stage. The second is to observe that an incumbent financier that imposes a hard budget constraint can credibly require a high initial repayment. The reason is that the entrepreneur faces a credible termination threat, since he cannot get continuation from his incumbent (given the hard budget constraint) or from a rival (given he already owes too much to the incumbent). Thus, the entrepreneur has incentive to work even if a lot of his output goes to his financier when his project succeeds, because he wants to avoid termination when the project does not succeed.

These results highlight the link between the entrepreneur's continuation value u_1 and his first-stage repayment R_1. In particular, Proposition 2 shows that whether the entrepreneur is captive depends on R_1 (inequality (10)). But observe that R_1 depends on the entrepreneur's continuation value u_1 via his first-stage IC (inequality (8)) and that u_1, in turn, depends on whether he is captive in the first place. Hence, finding u_1 is a fixed point problem. Solving it gives the following result.

Lemma 1. *Suppose*

$$K_1 > (p - \Delta)\frac{B}{\Delta}. \tag{12}$$

If an entrepreneur has a hard budget constraint (inequality (9) is violated), then his continuation value is $u_1 = 0$.

If he has a soft budget constraint (inequality (9) is satisfied), his continuation value is

$$u_1 = \begin{cases} \max\left\{p\frac{b}{\Delta}, \hat{Q}B\right\} & \text{if } \frac{b}{B} \le \frac{1}{\Delta} - \frac{1}{p} \text{ and } \hat{Q} \le \frac{p}{\Delta}\left(\frac{1}{\Delta} - \frac{1}{p}\right), \\ \hat{Q}p\frac{B}{\Delta} & \text{if } \frac{b}{B} \le \frac{1}{\Delta} - \frac{1}{p} \text{ and } \hat{Q} > \frac{p}{\Delta}\left(\frac{1}{\Delta} - \frac{1}{p}\right), \\ \frac{p}{\Delta}\max\left\{b, \hat{Q}B\right\} & \text{otherwise.} \end{cases} \tag{13}$$

[14] The idea that financiers can use a high cost of capital as a commitment device to withhold capital complements the intuition that financiers use their own leverage as a commitment device to prevent borrower opportunism. For example, in Diamond and Rajan (2001), the risk of depositor runs ensures a bank collects repayment from its borrowers, and in Axelson, Strömberg, and Weisbach (2009) intermediary leverage mitigates the conflict of interest between a private equity fund and its investors.

The relevant takeaway from the expression for u_1 above is that it is increasing in $\hat{Q}$: the entrepreneur's continuation value is higher when it is easier to get financing from a rival.

2.3 Assumptions

We now impose assumptions on the deep parameters.

We make two assumptions on parameters to distinguish the projects from each other.

Assumption 2. *Private benefits are high for HAC projects:*

$$B^A > 2B^a. \tag{14}$$

An entrepreneur's private benefits from shirking are higher with the HAC project. The specific restriction—that they are at least twice as high—implies that an entrepreneur gets more agency rent from the HAC project than from the LAC one, even if the LAC lasts twice as long (viz. two periods instead of one). As a result, he always prefers HAC projects. (Cf. Lemma 2.)

Assumption 3. *The initial investment cost is not too small or too large:*

$$(2-p)py^a - (1-p)K_1^a - 2p\frac{B^a}{\Delta} > K_0 > (2-p)py^A - (1-p)K_1^A - p\frac{B^A + b^A}{\Delta}. \tag{15}$$

This assumption says that the soft-budget-constraint problem is significantly more costly ex interim for the HAC project than the LAC project, because, basically, K_1^A is high relative to K_1^a. This makes it relatively unattractive to finance ex ante. The specific assumption ensures that a bank (which has cost of capital $\rho = 0$) will provide initial financing to an LAC-entrepreneur but not to an HAC-entrepreneur. (Cf. Lemma 4.)

We make the following assumption on non-banks' cost of capital.

Assumption 4. *Non-banks' cost of capital is not too small or too large:*

$$\frac{p\left(y^A - B^A/\Delta\right)}{K_0^A} \geq 1 + r > \frac{p\left(y^A - b^A/\Delta\right)}{K_1^A}. \tag{16}$$

This ensures that non-banks will provide financing to an HAC-entrepreneur at the first stage but not the second. (Cf. Lemma 3.)

We make the following assumption on the cost continuation financing.

Assumption 5. *The cost of continuation financing is not too small or too large: for HAC projects,*

$$p\left(y - \frac{b^A}{\Delta}\right) > K_1^A > (p-\Delta)\frac{B^A}{\Delta}, \tag{17}$$

and for LAC projects,

$$\max\left\{p^2\frac{b^a}{\Delta},(p-\Delta)\left(\frac{B^a}{\Delta}+p\frac{b^a}{\Delta}\right)\right\}\geq K_1^a>\frac{B^a}{(1+r)\Delta}\max\left\{p^2,(p-\Delta)(1+p)\right\}. \tag{18}$$

Condition (17) ensures that incumbent banks provide continuation financing to HAC entrepreneurs (cf. Lemma 4), but rivals do not (cf. Lemma 3). Condition (18) ensures that rival banks finance continuation of LAC entrepreneurs (cf. Lemma 4), but non-banks do not (cf. Lemma 5).

One example of a set of parameters satisfying all of these assumptions, as well as Assumption 1 and the hypothesis of Proposition 4 below, is as follows: $p=0.6,\Delta=0.4,r=5\%,K_0=60,y^A=175,K_1^A=80,B^A=22,b^A=15,y^a=110,K_1^a=7.5,B^a=9$, and $b^a=8.5$.

3 Equilibrium and Intermediation Variety

In this section, first, we characterize the subgame perfect equilibrium as a function of the level of competition Q and the mix of non-banks and banks φ and $1-\varphi$ and, second, we derive our main results on how this mix depends on the level of competition.

3.1 Equilibrium Characterization

The preliminary results and assumptions in the previous section allow us to characterize the equilibrium behavior of entrepreneurs, banks, and non-banks:

Proposition 3. (Equilibrium.) *For any level of competition Q among financiers and proportion φ of non-banks, the unique best responses are as follows:*

- *Entrepreneurs who meet non-banks choose HAC projects, their budget constraints are hard, and they are captive.*
- *Entrepreneurs who meet banks choose LAC projects, their budget constraints are soft, and they are not captive.*

This result says that the projects entrepreneurs choose depend on the kind of finance they have access to. All prefer HAC projects over LAC projects, because the former provide them higher agency rents (cf. Assumption 2). However, they can only invest in projects that they can finance, and different types of financiers are willing to finance different types of projects. In fact, only non-banks are willing to finance HAC projects. Anticipating this, entrepreneurs, who prefer HAC projects, which generate high agency rents, choose them when they meet to non-banks and choose LAC projects when they meet banks.

The reason that non-banks are willing to finance HAC projects is that they use their high cost of capital as a disciplining device. It allows them to commit not to refinance entrepreneurs, and thus to impose hard budget constraints (Proposition 1). This allows them to extract high initial repayments, thereby keeping entrepreneurs captive (Corollary 1). This disciplines entrepreneurs, who anticipate termination following failure, and hence provide effort in the first stage. Although this discipline decreases entrepreneurs' agency rents after projects are undertaken, they still welcome it, because it allows them to finance HAC projects in the first place.

The reason that banks, unlike non-banks, are unwilling to finance HAC projects is that, due to their low cost of capital, they cannot credibly commit not to refinance entrepreneurs (Proposition 1). Since entrepreneurs will always be able to refinance their projects, they have high continuation values u_1, making them costly to incentivize, especially with HAC projects. Indeed, this is so costly that a bank will not fund an HAC entrepreneur in the first place.

In our model only failure-intolerant financiers (non-banks) are willing to finance HAC projects. To the extent that HAC projects are likely to be innovative, this contrasts with the idea that failure tolerance fosters innovation (e.g., Manso (2011) and March (1991)). The reason is that even if failure tolerance is optimal for an individual entrepreneur, it could be prohibitively expensive for an external financier.[15]

The difference in financiers' cost of capital affects investment not only at the initial stage, but also at the continuation stage, when an entrepreneur could seek financing from a financier other than his incumbent. Rival banks are willing to provide continuation financing to LAC entrepreneurs, but rival non-banks are not. This leads to the following corollary:

Corollary 2. *The probability that an LAC entrepreneur finds continuation financing from a rival is the probability that he meets a bank:* $\hat{Q} = (1 - \varphi)Q$.

3.2 Intermediation Variety

We have established that entrepreneurs with access to non-banks choose HAC projects and those with access to banks choose LAC projects. But this does not answer the question: will there be a *mix* of banks and non-bank financiers in equilibrium? Or will financiers all prefer to be banks, benefiting from their lower cost of capital? Alternatively, will they all prefer to be non-banks, benefiting from their hard budget constraints? Moreover, does the mix of financiers in the market depend on the level of competition Q among them?

[15] Further, the agency problem in Manso (2011) is different from ours; he focuses on how to incentivize exploratory learning, not just effort.

To address these questions, we start by comparing the expected payoffs of banks and non-banks. Since financiers offer the contracts, they get the total surplus from a project less the agency rents they must surrender to incentivize entrepreneurs. In a meeting between an entrepreneur and a non-bank, the total surplus is the value of the HAC project, which can succeed in its first stage or not at all (given the hard budget constraint). We denote this by Σ^A. Noting that we need to discount the payoff by non-banks' cost of capital $1+\rho = 1+r$, we have

$$\Sigma^A := \frac{py^A}{1+r} - K_0. \tag{19}$$

Hence, since the non-bank meets an entrepreneur with probability q, the non-bank's expected payoff is

$$\text{non-bank's payoff } = q\left(\Sigma^A - p\frac{B^A}{\Delta}\right), \tag{20}$$

where the second term is the entrepreneur's expected rent $p(y - R_1)$.

In a meeting between an entrepreneur and a bank, the total surplus is the value of the LAC project. The project could succeed at either its first or its second stage (given the soft budget constraint). We denote this by Σ^a. Noting that we do not need to discount the payoff since the bank's cost of capital is $1+\rho = 1$, we have

$$\Sigma^a := py^a - K_0 + (1-p)\big(py^a - K_1^a\big). \tag{21}$$

Recalling that a bank meets an entrepreneur with probability q, a bank's expected payoff is

$$\text{bank's payoff } = q\left(\Sigma^a - \frac{p}{\Delta}B^a - u_1(\hat{Q})\right), \tag{22}$$

where the second term is the entrepreneur's expected rent $p(y-R_1)+(1-p)p(y-R_1-R_2)$. We have written the continuation value as $u_1(\hat{Q})$ to emphasize that u_1 depends on the probability the entrepreneur can get continuation financing from a rival financier.

Different types of financiers coexist if and only if their payoffs are equal in equilibrium, so a financier is indifferent between becoming a non-bank and a bank, or if:

$$q\left(\Sigma^A - p\frac{B^A}{\Delta}\right) = q\left(\Sigma^a - p\frac{B^a}{\Delta} - u_1(\hat{Q})\right). \tag{23}$$

This expression captures a key force in our model: an increase in competition among rival financiers, captured by an increase in $\hat{Q}$, increases the entrepreneur's continuation utility u_1 (Lemma 1) thereby exacerbating the bank's soft-budget-constraint problem. As a result, the bank must leave the entrepreneur a higher agency rent at the first stage (see the IC in equation (8)). This reduces the bank's profit on the RHS of equation (23). Thus, the more competitive the market is, the greater is the benefit to financiers from keeping entrepreneurs captive, making it more attractive to be a non-bank rather than a bank.

Re-writing, non-banks and banks co-exist if and only if the entrepreneur's continuation value is

$$u_1(\hat{Q}) = \frac{p}{\Delta}\left(B^A - B^a\right) - \left(\Sigma^A - \Sigma^a\right) =: u^*. \tag{24}$$

Given that, by Corollary 2, we know that the level of competition among rivals is $\hat{Q} = (1-\varphi)Q$. This expression for u^* (equation (24)) allows us to solve for the equilibrium mix of non-banks φ and banks $1-\varphi$ in the market as a function of the level of competition Q among all financiers. Indeed, rearranging, we see that if there is an interior mix of banks and non-banks in equilibrium, the proportion of non-banks is given by

$$\varphi = 1 - \frac{u_1^{-1}(u^*)}{Q} \tag{25}$$

(assuming the inverse of u_1 is well defined). And, there is indeed an interior mix of financiers as long as this expression is between zero and one. Since it can be less than zero, but never greater than one, we have:

$$\varphi = \max\left\{0\,,\,1 - \frac{u_1^{-1}(u^*)}{Q}\right\}. \tag{26}$$

Equation (26) implies that φ is an increasing function of Q, the level of competition among financiers. If Q is very low, φ is zero, indicating that no non-bank operates—the benefits of cheap capital (low ρ) outweigh the costs of soft-budget-constraint problems. As competition increases, banks' soft-budget-constraint problems become more severe, and some financiers become non-banks, helping to keep these problems at bay. But φ never reaches 1. Non-banks never take over the whole market, and banks provide some finance for all levels of competition Q. The proportion of non-banks approaches $1 - u_1^{-1}(u^*)$ in the perfect competition limit ($Q \to 1$), as depicted in Figure 3, and formalized in the next proposition.

Proposition 4. (Intermediation variety.) *Suppose that*

$$\Sigma^a - \frac{p}{\Delta}\left(B^a + b^a\right) > \Sigma^A - p\frac{B^A}{\Delta}. \tag{27}$$

1. *Non-banks are present only if competition among financiers Q is sufficiently high.*
2. *The proportion of non-banks is increasing in competition Q.*
3. *Non-banks never take over the entire market. Rather, banks provide a positive fraction of finance for all Q.*

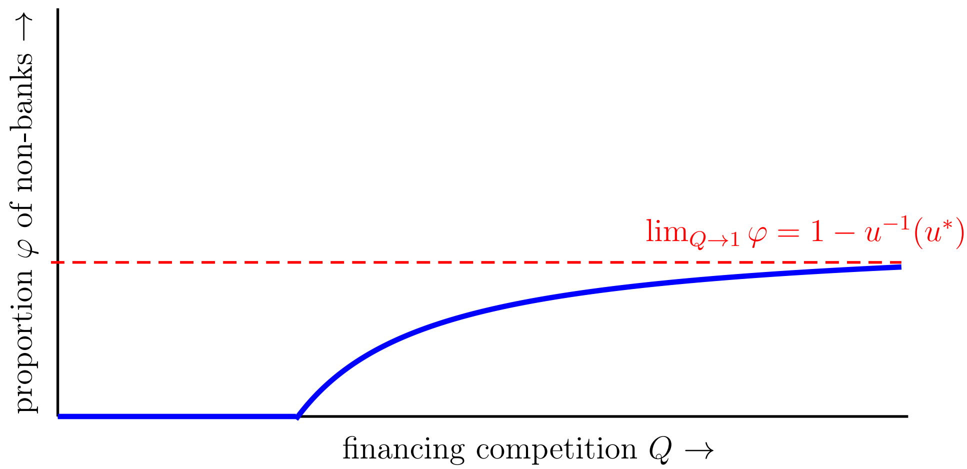

Figure 3: The proportion φ of non-banks in the market as a function of competition Q.

For low Q, all financiers become banks to take advantage of their funding cost advantage. But as competition increases, and there are more banks in the market, it becomes easier for entrepreneurs to find banks to finance their second-stage investments, i.e. $\hat{Q}$ goes up. As a result, entrepreneurs can extract more rents from their incumbent banks. Non-banks emerge in response, as the rents entrepreneurs can extract from them are limited. They keep entrepreneurs captive, and hence are always effectively monopolists, unaffected by competition. Still, non-banks do not provide all the finance for high competition. The reason is that non-bank entry attenuates the effect of competition on banks, because the higher φ is, the less sensitive $\hat{Q} = (1-\varphi)Q$ is to Q. Thus, if $\varphi \to 1$, then $\hat{Q} \to 0$: if no bank were to operate, then the probability an entrepreneur could find refinancing from a rival financier would go to zero. In this case, entrepreneurs would be effectively captive to banks. This would make banking desirable and induce banks to enter.

4 Discussion, Empirical Content, and Policy

In this section, we discuss our model's assumptions, empirical content, and policy implications.

4.1 Discussion

Banks' and non-banks' cost of capital. Underlying all of our results is the assumption that banks have a lower cost of capital than non-banks. It generates the high hurdle rate that non-banks apply to investments,[16] which in turn disciplines entrepreneurs, hardening their budget constraints.[17]

We stress above that banks' low cost of capital is likely due to their government guarantees and money-like deposits. But non-banks may have a higher cost of funding for other reasons as well. For example, unlike banks, non-banks such as VCs and PEs take on relatively few investments, so their undiversified positions and exposure to idiosyncratic risk could drive up their cost of capital.[18] Moreover, non-banks are likely to care more about upside payoffs, given that leverage and incentive distortions make their payoffs convex.[19] As a result, they may finance only entrepreneurs that still have high upside potential, which would also have the effect of hardening a soft budget constraint.[20] Finally, they are also likely to finance riskier investments and hence have higher probability of default themselves. This would drive up the rate they have to pay on their own financing to compensate their investors endogenously.

High- and low-agency-cost projects. For our results, the key distinction between the two types of projects is the severity of the soft-budget-constraint problem: compared to entrepreneurs with LAC projects, those with HAC projects are costly to incentivize and expensive to refinance. I.e. they have projects that have lower private benefits at each stage (B and b) and costs of continuation financing (K_1). We assume, however, that all projects have the same start-up costs (K_0), the same payoff given failure (zero), and the same success probabilities (p if entrepreneurs work and $p - \Delta$ otherwise). This simplifies the exposition, but it also means that we have to rely entirely on a few parameters to generate meaningful differences across projects.

In particular, Assumption 4 and Assumption 5 suggest that K_1^a should be "a lot" smaller than K_1^A. This could be reasonable, even taken literally. For example, refinancing innovative (HAC) projects could amount to starting over, whereas refinancing

[16]A few other papers show that VCs may impose high hurdle rates because the opportunity cost of their capital is high, even if their cost of capital is not; see Inderst, Mueller, and Münnich (2006), Jovanovic and Szentes (2013), and Khanna and Mathews (2017).

[17]Our model thus explains why some finance must be intermediated: non-banks' high cost of capital on the right-hand side of their balance sheets gives them the commitment power they need to make profitable investments on the left-hand side. See, e.g., Kashyap, Rajan, and Stein (2002) and Donaldson, Piacentino, and Thakor (2018) for other theories connecting intermediary assets and liabilities.

[18]E.g., Metrick and Yasuda (2010) find that the median VC fund expects to make only 20 investments over its lifetime and argue that "the expected number of investments plays an important role in driving the overall volatility of the fund portfolio, which in turn has a significant effect on the expected present value of revenue" (p. 2309).

[19]For example, finance companies lever their investments with bank debt limiting their downside risk (Chernenko, Erel, and Prilmeier (2018)), VCs want high upside payoffs to attract investor capital (Piacentino (2019)), and the general partners in PEs have contracts that reward them more on the upside (Axelson, Strömberg, and Weisbach (2009)).

[20]To capture this within our model, we need only to make the (reasonable) assumption that entrepreneurs' upside potential is higher at the first stage than the second. We abstract from this in the baseline for simplicity: it would amplify our results, but it is not necessary for them.

traditional (LAC) projects could be closer to minor upkeep. But it can also be taken as a stand-in for differences in other parameters. Most notably, to the extent that high agency costs capture innovative projects, it is likely that they actually do pay off zero in the event of failure (Hall and Woodward (2010)), whereas traditional projects are likely to have positive recovery value. Thus, the cost of continuation should be interpreted as only the new capital needed for an entrepreneur to continue a traditional project, which, net of the first period payoff, is likely to be relatively small for traditional projects compared to innovative ones.

In reality, HAC projects are likely to be riskier in the sense of having a lower success probability too. Our framework accommodates such heterogeneity without becoming intractable (although the equations do become significantly more complicated). We omit it only for simplicity.

Bilateral matching/Competition. We use a model of random bilateral meetings to embed a staged financing problem in market equilibrium. This is a useful set-up with precedent in the literature. It also allows us to do comparative statics on the level of competition, which is captured by an entrepreneur's probability of meeting a financier Q.

Financier's bargaining power. Throughout, we assume that financiers have the bargaining power when they negotiate contracts with entrepreneurs. This is useful from a modeling perspective, because it generates a division of surplus within a relatively classical principal-agent framework: entrepreneurs get agency rents and financiers get the remaining surplus (cf. equations (20) and (22)). However, it leaves open the question of how our results would change if entrepreneurs had some bargaining power, leaving them to propose lower repayments R_1. Here, we briefly discuss how this affects (or does not affect) each of our main results and explain why our main takeaways are robust.

- Decreasing R_1 does not affect our soft-budget-constraint result (Proposition 1). The reason is that an incumbent financier's willingness to provide continuation financing at the interim date does not depend on the amount the entrepreneur owes from the initial date, but only on the most he can promise to repay at the terminal date, as per the second-stage IC (per equation (6), no matter what R_1 is, the most he can promise to repay is $R_1 + R_2 = y - b/\Delta$).
- Decreasing R_1 could affect our endogenous captivity result (Proposition 2). The reason is that the entrepreneur is captive if and only if R_1 is sufficiently high (by equation (10)).

  However, the takeaway that HAC entrepreneurs are more likely to be captive will continue to hold as long as the repayment R_1 is sufficiently high with an HAC project relative to with an LAC project, which will be the case whenever agency rents tilt the division of surplus toward the entrepreneur. There are other realistic reasons that the financiers would be likely to require high repayments R_1 from HAC entrepreneurs, which we did not include in the baseline for simplicity; notably, HAC projects could be relatively risky ($p^A < p^a$) or require relatively large initial capital ($K_0^A > K_0^a$).

- Decreasing R_1 could affect our project-choice result (Proposition 3). The reason is that, with no bargaining power, entrepreneurs do not take the total surplus into account, and prefer the project that maximizes their agency rent. Thus, they choose the HAC whenever financing it is feasible.

  However, the takeaway that entrepreneurs choose HAC projects whenever they are feasible will continue to hold whenever they also have higher surplus, which is the case we have in mind to the extent that, for example, HAC projects are more innovative (we do not make this assumption in the baseline analysis only because it is not necessary for our results).

- Decreasing R_1 could affect our intermediation variety result (Proposition 4) to the extent that it affects our captivity result (described above). The reason is that the result relies on non-banks being monopolists over captive entrepreneurs, and hence is not affected by competition.

  However, the takeaway that the proportion of non-banks is increasing in competition will continue to hold as long as the endogenous captivity result does, which is likely to be the case for the reasons described above.

Incumbent's monitoring advantage. In our baseline model, we rely on the assumption that if an entrepreneur is refinanced by his incumbent financier, his private benefits are reduced from B to b, whereas if he is refinanced by a rival, they are not. This helps us to model imperfect competition. Even though financiers offer the contracts, the option to seek financing from a rival financier helps the entrepreneur to extract more surplus from his incumbent, because he can get higher private benefits/agency rents with the rival.

The assumption is intended to capture incumbent financiers' monitoring advantage, due, for example, to any propriety informational advantages they obtain in the course of their relationship with the entrepreneur (see, e.g., Rajan (1992)). We should stress, however, that when the entrepreneur gets financing from a rival at the second stage, his private benefits are not reduced, even if he has repayments to make to his incumbent from the first stage. This could be because information acquired during second-stage financing is complementary to that acquired during the first-stage relationship. Alternatively, it could be because the financier itself must have incentive to monitor, and does so only if it has a sufficiently large stake in the entrepreneur, as in

Holmström and Tirole (1997).[21] In particular, a financier must prefer to monitor at cost c, ensuring the entrepreneur works, and get its total repayment R_{tot} with probability p, than not to monitor, inducing the entrepreneur to shirk, and get R_{tot} with probability $p - \Delta$:

$$pR_{\text{tot}} - c \geq (p - \Delta)R_{\text{tot}} \tag{28}$$

or $R_{\text{tot}} \geq c/\Delta$ (which is (IC_{m}) on p. 672 of Holmström and Tirole (1997)). Thus, as long as $\max\{R_1, R_2\} < c/\Delta < R_1 + R_2$, a financier monitors if and only if it has provided finance at both the first and second stages (i.e. only if $R_{\text{tot}} = R_1 + R_2$).

4.2 Empirical Content

Banks in our model represent institutions that take deposits, and have a low cost of capital as a result. Non-banks could represent a variety of institutions that do not take deposits, but still compete with traditional banks to finance entrepreneurs. Salient examples are finance companies and venture capital firms. Others are private equity firms,[22] asset managers, and commercial mortgage banks. Given this, we now cite evidence for the facts at the start of the paper. Afterward, we discuss some other evidence consistent with our model.

Motivating facts.

(i) *Compared to banks, non-banks finance relatively high-agency-cost entrepreneurs.*

- Evidence on this prediction requires proxies for firms/entrepreneurs with high agency costs. Possible proxies include firms with low asset tangibility, high growth options, high asset specificity (see Gompers (1995)) as well as those that are young, are risky/have low credit quality, are innovative/R&D-intensive/high-tech, or have low current profitability.
- Given these proxies, see Chernenko, Erel, and Prilmeier (2018), Carey, Post, and Sharpe (1998), and Denis and Mihov (2003) on finance companies, Gompers (1995), Kortum and Lerner (2000), and Hellmann and Puri (2000) on VCs, and Lerner, Sorensen, and Strömberg (2011) on PEs.

(ii) *Compared to banks, non-banks charge entrepreneurs relatively high rates.*

- See Chernenko, Erel, and Prilmeier (2018), who document that non-bank-loans carry 190 basis points higher interest rates than bank loans.

[21] Using data on credit lines, Acharya et al. (2014) find empirical support for the predictions of this model of monitoring.

[22] Our model might not apply to LBOs, in which PEs often target low-risk firms. But it could apply to other branches of the PE business. Indeed, anecdotally, it seems PEs are increasingly competing with banks in the lending market. See, e.g., "The New Business Banker: A Private Equity Firm," *Wall Street Journal*, August 12, 2018 and "How the Biggest Private Equity Firms Became the New Banks," *Financial Times*, September 19, 2018.

- See Cochrane (2005), Hall and Woodward (2007), Korteweg and Sorensen (2010), and Korteweg and Nagel (2016) who document abnormal returns on VCs' portfolio investments.

(iii) and (iv) *Compared to banks, non-banks have relatively short-term relationships and are relatively intolerant of failure.*

- See Kerr, Nanda, and Rhodes-Kropf (2014), who find that VC backed firms are likely to be shut down relatively early, despite high upside potential. Indeed, Gompers and Lerner (2001) say that "[s]taged capital infusion may be the most potent control mechanism a VC can employ" (p. 155). And, Guler (2018) identifies VCs' ability to terminate failing investments as a primary driver of their success, suggesting it could be on par with picking winning entrepreneurs in the first place. Chernenko, Erel, and Prilmeier (2018) find that many non-bank lenders, including hedge funds and other asset managers, use relatively short-maturity loans, which they suggest discipline borrowers.[23]
- See Sahlman (1990), who describes how VCs, unlike banks, finance firms with the intent to exit, likely in an IPO, and to terminate otherwise.
- See Gompers (1995), who finds that high-agency-cost entrepreneurs are associated with shorter financing duration.

(v) *Compared to banks, non-banks exist in relatively competitive financial markets.*

- See, e.g., Chernenko, Erel, and Prilmeier (2018), who show that firms are more likely to get financed by non-banks when banking is more concentrated.
- See also Boyd and Gertler (1994), who attribute the decline in the share of C&I loans provided by banks partly to increases in competition following the deregulation of the 1980s. Neuhann and Saidi (2016) also find that "[b]ank deregulation thus facilitated the entry of non-bank intermediaries into the market for corporate credit."
- To the extent that increasing competition decreases market values, see Irani et al. (2020), who find that low bank capitalization leads to non-bank entry. Similarly, the IMF (2016) attributes the rise in non-bank finance in part to weak bank balance sheets. It also underscores that non-banks are less prevalent in less developed credit markets, which are likely have the highest impediments to competition.

(vi) *Compared to banks, non-banks are relatively scarce.*

[23]They find that insurance companies are an exemption; they lend long-term. To the extent that insurance companies benefit from a low cost of capital (since their liabilities are their insurance policies), this is arguably consistent with our theory.

- See, e.g., Puri and Zarutskie (2012), who find that only a fraction of a percent of new companies are VC funded in US Census data.
- See also Chernenko, Erel, and Prilmeier (2018), who find that non-bank finance constitutes less than a third of the loans in their sample of mid-market firms.

(vii) *Compared to bank-financed entrepreneurs, non-bank-financed entrepreneurs do relatively high-agency-cost projects.*

- This is the counterpart of (i) above. See the evidence cited there.
- See also the anecdotal evidence below that high-agency-cost entrepreneurs actively seek out non-bank finance.

(viii) *Compared to bank-financed entrepreneurs, non-bank-financed entrepreneurs are relatively unlikely to obtain financing from other financiers.*

- On VC, see Berger and Schaek (2011), who show that entrepreneurial firms substitute venture capital for multiple banking relationships. See also Barry and Mihov (2015) and Rin, Hellmann, and Puri (2013), who report that entrepreneurs who get more VC financing get less bank financing and that VC-backed companies are less likely to have multiple banking relationships.
- Anecdotally, companies borrowing from non-banks often do not have access to bank credit ("Bain and BlackRock expand their Asia private credit businesses," *Financial Times*, June 14 2017). See Lim, Minton, and Weisbach (2014) for evidence in the context of leverage loans.

Other evidence. One feature that distinguishes our model from other theories of entrepreneurial finance is that in our model entrepreneurs choose their projects in response to the kind of finance they have access to, not the other way around (see Proposition 3). In the model, this happens in an extreme way: an entrepreneur is matched with one financier that it can get finance from. However, it could reflect something milder, such as a choice made before matching with financiers, but in anticipation of the pool of available financiers—e.g., entrepreneurs located in Silicon Valley may have more access to VC finance than those elsewhere—or a choice to "tilt" the project in some direction made during negotiations with a financier. Direct evidence on this prediction is lacking, probably because it requires information about the set of projects entrepreneurs have available but do not undertake. However, it resonates with indirect evidence:

(ix) *Access to (non-bank) finance determines project choice for high-agency-cost entrepreneurs.*[24]

[24]It is critical for our results that project choices happen after entrepreneurs and financiers meet.

- See Sorenson and Stuart (2001), who find that entrepreneurs further from VCs are more likely to be denied financing; see Blanchflower and Oswald (1998), who find that one of the main reasons people choose not to pursue entrepreneurship is that they have limited access to financing; and see Samila and Sorenson (2011), who find that an increase in the supply of venture capital makes people more likely to engage in entrepreneurship.
- Anecdotally, access not only to financing, but to the right *type* of financing, is a first-order consideration for entrepreneurs. For example, access to VC financing is among the most-cited reasons why entrepreneurs decide to headquarter in the Bay Area (e.g., Cohan (2013) and Wessel (2013)); similarly, Chen et al. (2010) find that location is related to VC outcomes.
- See the Kauffman survey for evidence that high-agency-cost entrepreneurs (proxied by "insufficient collateral," as discussed above) must "'take what they can get' rather than the financing that would be the best fit for their needs." The survey also stresses that entrepreneurs do not apply for financing when they fear being denied. This is in line with our model, in which entrepreneurs alter their project choices to avoid being denied finance.
- More generally, see Hellmann and Puri (2000), Kerr, Lerner, and Schoar (2011), and Lerner et al. (2015), who find that access to finance is a driver of innovation.

In our baseline set-up, an increase in competition among financiers leads to more non-bank entry and therefore more entrepreneurs choosing high-agency-cost projects in anticipation of non-bank finance. However, as we show in an extension (Section 5.1), this specific prediction of our model actually depends on our parametric assumptions. The reason is that, in general, when entrepreneurs choose projects, they face a trade-off. With the HAC project, they get high agency rents, but are captive. With the LAC project, they get low agency rents, but are not captive. So far, we have focused on the case in which the HAC project offers a significant increase in rents, and thus is preferred by entrepreneurs. However, if the HAC project offered only a modest increase in rents, then entrepreneurs could prefer the LAC project, even if the HAC project is efficient, in the sense that the payoff y^A is sufficiently high relative to y^a. Thus, our model suggests that the effect of competition on project choice is ambiguous. To the extent that innovative projects proxy for high-agency-cost projects, the empirical evidence is mixed as well:

(x) *Real-sector innovation can be increasing or decreasing in banking competition.*

- See Chava et al. (2013), and Mao and Wang (2018).[25]

[25] To the extent that banks finance with debt and non-banks with equity, see also Hsu, Tian, and Xu

- Hombert and Matray (2016), Cornaggia et al. (2015), and (under the assumption that innovation is relatively risky) Kaviani and Maleki (2018) provide evidence that real-sector innovation declines with banking competition.

In addition to suggesting that banks provide longer-term financing to entrepreneurs than non-banks, our model could potentially illuminate other details of contracts.

(xi) *Contract terms.*

- In our model, banks, with their soft budget constraints, provide refinancing at favorable terms. See Degryse and Cayseele (2000) for evidence that contract terms deteriorate with the duration of financing relationships.
- To the extant that banks, unable to rely on a credible termination threat, could substitute with contractual terms, our model suggests that banks should use relatively more covenants. See Chernenko, Erel, and Prilmeier (2018) for evidence consistent with this.

Although our banks and non-banks are identical in every way except for their cost of capital, our results are consistent with classic findings about the unique value of banking relationships, stressed in the relationship-banking literature. In particular, ex interim, entrepreneurs value their relationships with banks, which provide them continuation finance after failure, but not with non-banks, which terminate. Unlike in the literature, however, this difference does not depend on assumed differences in information, monitoring ability, or horizon (i.e. myopia or lack thereof).

(xii) *Value of banking relationships.*

- See, e.g., Degryse and Ongena (2005) and Nguyen (2019), who find that bank branch closings harm local borrowers.

4.3 Policy

Our model stresses that the projects and technologies developed by entrepreneurs depend on the availability of financiers to fund them. This could have implications for policy.[26]

(2014), who find equity market development increases innovation, whereas debt market development seems to decreas it. Keep in mind, however, that debt and equity are theoretically equivalent in our set-up.

[26] Although we comment on specific policy objectives, we refrain from giving a formal definition of efficiency. There are two main reasons for this. (i) With imperfect markets (bilateral matching) and heterogeneous agents (different costs of capital), there is not a clear way to define the discount rate that determines whether one project is better than another in an NPV sense. (ii) With financiers' funding cost difference taken in reduced form, there is not a clear way to define aggregate welfare. To do so, we would have to take a stand on where the difference comes from, and close the model more fully. We choose not to do this, given the cost of capital difference could reflect a variety things (Section 4.1). Just for example, banks' low cost of capital

For example, more innovative or productive projects may also have high agency costs, and therefore require non-bank financing.

In this case, a policy maker could want to foster non-bank entry to encourage entrepreneurs to develop efficient technologies. As such, he could consider subsidizing non-bank funding, perhaps trying to level the playing field with banks, which already benefit from government guarantees. Our model suggests, however, that such policies could backfire. The reason is that decreasing non-banks' cost of capital could undermine the credibility of their termination threat, making them unable to finance HAC entrepreneurs.

An alternative way to encourage non-bank entry would be to tax banks, making non-banking relatively attractive. Tightening bank regulation could also have a similar effect and lead to an increase in non-banking, as Buchak et al. (2018) and Irani et al. (2020) document. Our analysis suggests that such regulatory arbitrage, typically cast in a negative light, could have a bright side.

Finally, our model speaks to the unintended consequences of deposit insurance, in so far as the deposit insurance subsidies lower banks' cost of capital and undermine their threat to withhold capital from entrepreneurs. Our analysis suggests that banks' contributions to deposit insurance schemes should be designed to minimize this distortion.

5 Extensions

In this section, we describe three extensions. First, we relax the assumption that banks do not fund HAC projects. Then, we allow for so-called "congestion externalities." Finally, we allow for the possibility that HAC projects are scarce.

5.1 Entrepreneurs Choose Not to Innovate

So far, entrepreneurs who met banks chose LAC projects because they knew they could not get funding for HAC projects. We now turn to another reason that they might choose not to do HAC projects: to avoid being captive. Specifically, if they anticipate being captive if they do HAC projects, but not LAC ones, they could prefer to do LAC projects, which allow them to refinance at better terms. This cannot happen given the parameter assumptions of the baseline model, because we assume that HAC projects

could reflect a social purpose played by safe deposits, and these deposits could be safe in part because they are backed by LAC projects. In this case, investing in LAC projects need not be inefficient. In contrast, the low cost of capital could reflect fiscal backing by the government. In this case, investing in LAC projects likely would be inefficient.

are viable only if financiers can impose hard budget constraints. However, it can happen for other parameters.

Proposition 5. (Entrepreneurs choose not to innovate.) *Suppose the assumptions in the baseline model hold, except that condition (64) for banks ($\rho = 0$) holds for HAC but not LAC entrepreneurs (which implies HAC but not LAC entrepreneurs are captive to banks) and that*

$$K_0^A < py^A + (1-p)\Big(py^A - K_1^A\Big) - 2p\frac{B^A}{\Delta} \tag{29}$$

(which implies that HAC projects are viable for banks).

Entrepreneurs matched with banks choose LAC projects (possibly inefficiently) if and only if interbank competition $\hat{Q}$ is sufficiently high.

Intuitively, when entrepreneurs choose projects, they face a trade-off. With the HAC project, they get high private benefits, but are captive; with the LAC project, they get low private benefits, but are not captive. Thus, they may choose the LAC project, even if the HAC project is efficient, in the sense that the payoff y^A is sufficiently high relative to y^a. This points to another way that non-banks' high cost of capital can discipline entrepreneurs. Not only does it allow non-banks to commit to deny second-stage financing, and harden entrepreneurs' soft budget constraints, but it also allows them to commit to deny financing to LAC projects, which forces entrepreneurs to do HAC projects. Consequently, entrepreneurs with access to banks choose LAC projects, and those with access to non-banks choose HAC projects.

5.2 Congestion

Here we show that our results are robust to, and sometimes amplified by, the possibility that similar financiers compete for the same entrepreneurs, in the sense that the probability that a bank meets an entrepreneur is decreasing in the number of other banks operating and the probability that a non-bank meets an entrepreneur is decreasing in the number of other non-banks. This captures so-called congestion externalities, which are a hallmark of models of markets in which trading/search frictions can make it hard to find a counterparty, e.g., the Mortensen and Pissarides (1994) model of labor markets, the Duffie, Gârleanu, and Pedersen (2005) model of OTC asset markets, and the Inderst and Mueller (2004) model of venture capital markets.[27] To include congestions externalities, we suppose that banks and non-banks meet entrepreneurs with the

[27] Such congestion externalities can also be present in our baseline set-up (although they need not be). However, unlike in this extension, they affect all financiers the same way; i.e. when a non-bank enters, it imposes the same externalities on banks as on other non-banks.

"telephone" probabilities (Stevens (2007)):[28]

$$q_{\text{nb}} := \frac{1}{1+\frac{\varphi}{1-Q}} \quad \text{and} \quad q_b := \frac{1}{1+\frac{1-\varphi}{1-Q}}. \tag{30}$$

Now, financiers' indifference condition reads

$$q_{\text{nb}}\left(\Sigma^A - p\frac{B^A}{\Delta}\right) = q_b\left(\Sigma^a - p\frac{B^a}{\Delta} - u_1(\hat{Q})\right). \tag{31}$$

which is just equation (23) with non-banks' and banks' matching probability q replaced by q_{nb} and q_b. Following the analysis in Section 3.2, we can rearrange equation (31) to write

$$u_1(\hat{Q}) = \Sigma^a - p\frac{B^a}{\Delta} - \frac{q_{\text{nb}}}{q_b}\left(\Sigma^A - p\frac{B^A}{\Delta}\right) \tag{32}$$

$$= u^* + \left(1 - \frac{q_{\text{nb}}}{q_b}\right)\left(\Sigma^A - p\frac{B^A}{\Delta}\right). \tag{33}$$

Solving for φ and comparing the perfect competition limit ($Q \to 1$) to that in the baseline model gives the next result:

Proposition 6. (Intermediation variety with congestion.) *Suppose that the conditions of Proposition 4 hold, that there is congestion within banks and non-banks as specified in equation (30), and that* $\frac{b}{B} \leq \frac{1}{\Delta} - \frac{1}{p}$. *In the perfect competition limit* ($Q \to 1$), *the proportion of non-banks is given by*

$$\varphi_c \to \frac{1}{2}\left(1 - \frac{u^*}{pB^a/\Delta} - 2\beta + \sqrt{\left(1 - \frac{u^*}{pB^a/\Delta} - 2\beta\right)^2 + 4\beta}\right). \tag{34}$$

where

$$\beta := \frac{\Sigma^A - p\frac{B^A}{\Delta}}{pB^a/\Delta}, \tag{35}$$

(and the expression in equation (34) is well defined between zero and one).

The limiting proportion of non-banks is higher than in the baseline model if and only if it is less than half in the baseline model.

The result says that congestion in each market works as an additional equilibrating force, bringing the limiting proportion of non-banks closer to a half. The reason is that congestion pulls against the thin market, be it the market of banks or of non-banks.

[28] These probabilities would more commonly be written as a function of the number of financiers that enter. To economize on notation, we write everything in terms of the probability Q that an entrepreneur meets a financier, rather than introducing a notation for this number (which, given telephone matching probabilities with parameters equal to one, is just $\frac{1}{1-Q}$).

5.3 Scarcity of Innovative Projects

Here we show that our results are robust to, and indeed amplified by, the possibility that there could be relatively few truly innovative ideas available. With the idea that HAC projects correspond to innovation, we capture this with the assumption that the total supply of innovative projects is at most $S^A < 1$. We maintain the assumption that entrepreneurs are ex ante identical, but we suppose that if there are $E^A > S^A$ innovative entrepreneurs, each gets a viable project with probability S^A/E^A, and otherwise gets zero. If $E^A \leq S^A$, they all get viable projects, as in the baseline set-up. Thus, if there are only a few financiers in the market (low Q) our assumption here that innovative projects are limited does not affect our analysis above, since few innovative projects are funded anyway. For high Q, however, becoming a non-bank becomes less attractive, so there are fewer non-banks. This strengthens our result.

To see why, observe that financiers' indifference condition now reads

$$\min\left\{1, \frac{S^A}{E^A}\right\}\left(\Sigma^A - p\frac{B^A}{\Delta}\right) = \Sigma^a - p\frac{B^a}{\Delta} - u_1(\hat{Q}), \tag{36}$$

which is just the baseline indifference condition in equation (23) with non-banks' payoff multiplied by the probability of successful innovation, i.e. by $\min\left\{1, S^A/E^A\right\}$.[29] Following the analysis in Section 3.2, we can rearrange equation (31) to write

$$u_1(\hat{Q}) = \Sigma^a - p\frac{B^a}{\Delta} - \min\left\{1, \frac{S^A}{E^A}\right\}\left(\Sigma^A - p\frac{B^A}{\Delta}\right) \tag{37}$$

$$= u^* + \left(1 - \min\left\{1, \frac{S^A}{E^A}\right\}\right)\left(\Sigma^A - p\frac{B^A}{\Delta}\right). \tag{38}$$

Comparing the perfect competition limit ($Q \to 1$) to that in the baseline model gives the next result:

Proposition 7. (Intermediation variety with scarce innovative projects.) *Suppose that the conditions of Proposition 4 hold and that there is a limited supply S^A of innovative projects, assumed not to be too small. The proportion of non-banks φ_s is smaller than it is in the baseline model with elastic supply.*

Intuitively, a scarce supply of innovative projects makes it less attractive to become a non-bank, since a non-bank may end up with an entrepreneur lacking a viable idea. Hence, fewer financiers become non-banks.

[29] We should point out that the number of innovative entrepreneurs, E^A, is itself endogenous. In fact, it is just equal to the probability an entrepreneur meets a non-bank, $E^A = \varphi Q$, given entrepreneurs innovate if and only if they meet non-banks. We do not substitute it here, however, because it is not necessary for our result below.

6 Conclusion

We have developed an equilibrium model in which banks and non-banks co-exist even though banks have a lower funding cost than non-banks. This apparent disadvantage that non-banks face becomes an advantage in dealing with the soft-budget-constraint problem vis-à-vis borrowers. Because of their higher funding cost, non-banks are able to more credibly threaten not to continue financing the entrepreneur, thereby enabling them to deal more effectively than banks with borrowers when agency costs are high. Unlike previous theories, entrepreneurs' project choices depend on the type of financiers they have access to. Entrepreneurs with access to banks choose low-agency-cost projects and those with access to non-banks choose high-agency-cost projects, thus providing market segmentation with intermediation variety. The theory is consistent with a number of stylized facts about bank and non-bank financing, and it gives a new perspective on some policies.

A Proofs

A.1 Proof of Proposition 1

An entrepreneur's budget constraint is soft if and only if his incumbent financier is willing to finance continuation. First, we derive a necessary condition for the incumbent to finance continuation. Then, we explain that this condition is also sufficient.

There are two cases, either (i) the financier offers the maximum R_2 that satisfies the entrepreneur's second-stage IC, and the entrepreneur works; or (ii) it offers the maximum R_2 that the entrepreneur can feasibly repay, i.e. that satisfies the "feasibility constraint,"

$$R_1 + R_2 \leq y, \tag{39}$$

and the entrepreneur shirks. The incumbent imposes an SBC if and only if his second-stage PC is satisfied in either of these cases. We consider them in turn.

- **Case (i): Entrepreneur works:** The maximum repayment R_2 that satisfies the entrepreneurs' second-stage IC (inequality (6)) with $\beta = b$ solves

$$R_1 + R_2 = y - \frac{b}{\Delta}. \tag{40}$$

  The incumbent financier's expected payoff is

$$\begin{aligned} \Pi_2^{\text{work}} &:= p(R_1 + R_2) \\ &= p\left(y - \frac{b}{\Delta}\right). \end{aligned} \tag{41}$$

- **Case (ii): Entrepreneur shirks:** The maximum repayment that the entrepreneur can make that satisfies his feasibility constraint solves

$$R_1 + R_2 = y, \tag{42}$$

  and he does not work. Hence, in this case, the financier's expected payoff is

$$\begin{aligned} \Pi_2^{\text{shirk}} &: = (p - \Delta)(R_1 + R_2) \\ &= (p - \Delta)y. \end{aligned} \tag{43}$$

Observe from the expressions above that it is always the case that the financier prefers to incentivize work:

$$\Pi_2^{\text{work}} \geq \Pi_2^{\text{shirk}} \iff \Delta^2 y \geq pb, \tag{44}$$

which is implied by Assumption 1 (since $B > b$). Hence, a necessary condition for the creditor to provide continuation financing is that $\Pi_2^{\text{work}} \geq (1+\rho)K_1$, which is condition (9) in the lemma.

This condition need not be sufficient. Although it holds for the maximum repayments—the incumbent can extract these if the entrepreneur is captive—it need not hold for lower repayments, which the incumbent could potentially have to offer compete with a rival financier with a lower discount rate. However, it is indeed sufficient. The reason is that the only way that an entrepreneur is not captive is if a rival is willing to finance him. But, in this case, the incumbent is willing to finance too (this follows from Corollary 1). □

A.2 Proof of Proposition 2

An entrepreneur is not captive if and only if a rival financier is willing to provide continuation financing.

As in the proof of Proposition 1, there are two cases, either (i) the financier offers the maximum R_2 that satisfies the entrepreneur's second-stage IC constraint, and the entrepreneur works, or (ii) it offers the maximum R_2 that satisfies the entrepreneur's feasibility constraint, and the entrepreneur shirks. The entrepreneur is captive if and only if the rival financier's second-stage PC is satisfied in neither of these cases. We consider them in turn.

- **Case (i): Entrepreneur works:** If the rival financier incentivizes the entrepreneur to work, it offers the repayment $\hat{R}_2$ so that his second-stage IC binds (inequality (6)) with $\beta = B$.

$$R_1 + \hat{R}_2 = y - \frac{B}{\Delta}. \tag{45}$$

  The rival financier's payoff in this case is

$$\hat{\Pi}_2^{\text{work}} := p\hat{R}_2 \tag{46}$$

$$= p\left(y - \frac{B}{\Delta} - R_1\right). \tag{47}$$

- **Case (ii): Entrepreneur shirks:** If the rival financier extracts the highest repayment ex post, it offers $\hat{R}_2$ so that the entrepreneur's feasibility constraint binds (inequality (39)), or

$$R_1 + \hat{R}_2 = y. \tag{48}$$

The rival financier's payoff in this case is

$$\begin{aligned}\hat{\Pi}_2^{\text{shirk}} &:= (p-\Delta)\hat{R}_2 \\ &= (p-\Delta)\,(y-R_1)\,.\end{aligned} \tag{49}$$

Substituting from the above into into the rival's PC (inequality (3)), we see that the entrepreneur is captive if and only if

$$\begin{aligned}(1+\rho)K_1 &> \max\left\{\hat{\Pi}_2^{\text{work}}\,,\,\hat{\Pi}_2^{\text{shirk}}\right\} \\ &= \max\left\{p\left(y-\frac{B}{\Delta}-R_1\right)\,,\,(p-\Delta)\,(y-R_1)\right\}.\end{aligned} \tag{50}$$

This is gives the condition in the proposition, which holds if and only if R_1 is sufficiently high.

Finally, we make a side note that will be useful later. First, we note that by comparing the cases above we get a condition for the rival to prefer to incentivize work:

$$\hat{\Pi}_2^{\text{work}} \geq \hat{\Pi}_2^{\text{shirk}} \iff \Delta^2(y-R_1) \geq pB. \tag{51}$$

□

A.3 Proof of Corollary 1

The initial financier wants to set the highest possible R_1, subject to the entrepreneur's first-stage IC (which is satisfied by Assumption 1; see Appendix B). The maximum R_1 he can set is $R_1^{\max} := y - B/\Delta$, which comes from the binding IC with $u_1 = 0$. The entrepreneur is captive (and hence $u_1 = 0$ is consistent with the equilibrium) if inequality (10) is satisfied given $R_1 = R_1^{\max}$ or

$$K_1 > \max\left\{0, (p-\Delta)\frac{B}{\Delta}\right\}, \tag{52}$$

which is the condition in the corollary. □

A.4 Proof of Lemma 1

Solving for u_1 involves considering a number of cases, corresponding to whether the entrepreneur faces an HBC/SBC, is captive/not, and whether he works/shirks. The proof involves going through these cases.

If the entrepreneur faces an HBC, under the condition in the lemma, we know that he is also captive (by Corollary 1). Hence, his continuation value is zero.

If he faces an SBC, there are two cases, each with two subcases, as follows: he can be captive to his incumbent financier or not, and, in each case, his IC can be binding (if the financier incentivizes him to work) or not (if it does not).

1. **Case (i): Entrepreneur captive:** In this case, the entrepreneur's incumbent financier will finance him but a rival will not. Given Assumption 1, the incumbent will always make the entrepreneur's second-stage IC bind, i.e. $R_1 + R_2 = y - b/\Delta$ (cf. the proof of Proposition 1). This leaves the entrepreneur an agency rent of:

$$\begin{aligned} u_1 &= p(y - R_1 - R_2) \\ &= p\frac{b}{\Delta}. \end{aligned} \tag{53}$$

2. **Case (ii): Entrepreneur not captive:** If the entrepreneur is not captive, then the incumbent offers R_2 so that he prefers not to look for a rival. Hence, the entrepreneur's payoff is the greater of (i) the payoff he gets from his incumbent with a binding IC (as in the case of captivity; see equation (53)) and (ii) the expected payoff he would get from looking for a rival (recall that he gets higher private benefits with a rival, which cannot monitor).

   Thus, to find his continuation value, we need to compute his payoff if he meets a rival and then multiply it by the probability that the rival funds him, $\hat{Q}$. As in the previous results, there are two subcases: either (a) the rival offers the maximum $\hat{R}_2$ that satisfies the entrepreneurs' second-stage IC (and the entrepreneur works) or (b) it offers the maximum $\hat{R}_2$ that satisfies his feasibility constraint (and the entrepreneur shirks). We know from the proof of Proposition 2, that we are in subcase (a) if inequality (51) is satisfied and subcase (b) otherwise. We now compute u_1 in each subcase:

   - **Subcase (a): Entrepreneur works:** In this case, the entrepreneur's total repayment $R_1 + \hat{R}_2$ is given by equation (45). His expected payoff from looking for a rival financier is

$$\begin{aligned} u_1 &= \hat{Q}p\left(y - R_1 - \hat{R}_2\right) \\ &= \hat{Q}p\frac{B}{\Delta}. \end{aligned} \tag{54}$$

   - **Subcase (b): Entrepreneur shirks:** In this case, the entrepreneur's repayment is such that his feasibility constraint (inequality (39)) binds and he gets only his private benefits. His expected payoff from looking for a rival is

$$u_1 = \hat{Q}B. \tag{55}$$

Using the above and the condition for the rival to incentivize work (inequality (51)), we can write u_1 for the not-captive entrepreneur as:

$$u_1 = \begin{cases} \max\left\{p\frac{b}{\Delta}, \hat{Q}B\right\} & \text{if } \Delta^2(y - R_1) < pB, \\ \max\left\{p\frac{b}{\Delta}, \hat{Q}p\frac{B}{\Delta}\right\} & \text{otherwise.} \end{cases} \tag{56}$$

Note that u_1 (on the LHS) depends on R_1 (on the RHS), which, in turn, depends on u_1, so this equation embeds a fixed-point problem, which could have multiple solutions. In the case that u_1 above is multi-valued, it takes the smallest value. The reason is that the initial financier offers the highest R_1 it credibly can and in so doing "chooses" the lowest possible u_1.

Recall that the first-stage IC (inequality (8)) is satisfied by Assumption 1 (see Appendix B). Given the initial financier offers the highest possible R_1, the IC will bind (inequality (8)). This gives an expression for R_1 in terms of u_1:

$$R_1 = y - \frac{B}{\Delta} - u_1. \tag{57}$$

Substituting this into the expression for u_1 above, we have that

$$u_1 = \begin{cases} \max\left\{p\frac{b}{\Delta}, \hat{Q}B\right\} & \text{if } \max\left\{p\frac{b}{\Delta}, \hat{Q}B\right\} \leq \frac{B}{\Delta}\left(\frac{p}{\Delta} - 1\right), \\ \max\left\{p\frac{b}{\Delta}, \hat{Q}p\frac{B}{\Delta}\right\} & \text{if } \max\left\{p\frac{b}{\Delta}, \hat{Q}p\frac{B}{\Delta}\right\} \geq \frac{B}{\Delta}\left(\frac{p}{\Delta} - 1\right). \end{cases} \tag{58}$$

The remainder of the proof consists of simplifying the expression above case by case. There are three cases, which we write as (i) low, (ii) medium, or (iii) high pb/Δ:

- **Case (i):** $pb/\Delta \leq \hat{Q}B$**.** This corresponds to $\hat{Q} \geq \frac{p}{\Delta}\frac{b}{B}$. In this case,

$$u_1 = \begin{cases} \hat{Q}B & \text{if } \hat{Q} \leq \frac{p}{\Delta}\left(\frac{1}{\Delta} - \frac{1}{p}\right), \\ \hat{Q}p\frac{B}{\Delta} & \text{if } \hat{Q} \geq \frac{1}{\Delta} - \frac{1}{p}. \end{cases} \tag{59}$$

  Comparing the thresholds above, we see that, given $\frac{1}{\Delta} - \frac{1}{p} < \frac{p}{\Delta}\left(\frac{1}{\Delta} - \frac{1}{p}\right)$ (because $p > \Delta$), there is a non-empty interval in which the above is multi-valued. In this case, it takes the smaller value, namely $\hat{Q}B$. Hence,

$$u_1 = \begin{cases} \hat{Q}B & \text{if } \hat{Q} \leq \frac{p}{\Delta}\left(\frac{1}{\Delta} - \frac{1}{p}\right), \\ \hat{Q}p\frac{B}{\Delta} & \text{if } \hat{Q} > \frac{p}{\Delta}\left(\frac{1}{\Delta} - \frac{1}{p}\right). \end{cases} \tag{60}$$

- **Case (ii): $\hat{Q}B < pb/\Delta \leq \hat{Q}pB/\Delta$.** This corresponds to $b/B \leq \hat{Q} < pb/(\Delta B)$. In this case,

$$u_1 = \begin{cases} p\frac{b}{\Delta} & \text{if } \frac{b}{B} \leq \frac{1}{\Delta} - \frac{1}{p}, \\ \hat{Q}\,p\frac{B}{\Delta} & \text{if } \hat{Q} \geq \frac{1}{\Delta} - \frac{1}{p}. \end{cases} \tag{61}$$

Comparing the thresholds above, we see that it always has at least one well-defined value (given the conditions of Case (ii)), but it can be multi-valued. In this case, it takes the smaller value, namely pb/Δ:

$$u_1 = \begin{cases} p\frac{b}{\Delta} & \text{if } \frac{b}{B} \leq \frac{1}{\Delta} - \frac{1}{p}, \\ \hat{Q}p\frac{B}{\Delta} & \text{if } \frac{b}{B} > \frac{1}{\Delta} - \frac{1}{p}. \end{cases} \tag{62}$$

- **Case (iii): $pb/\Delta > \hat{Q}pB/\Delta$.** This corresponds to $\hat{Q} < b/B$. In this case,

$$u_1 = p\frac{b}{\Delta}. \tag{63}$$

Collecting the cases above gives the expression in the lemma.[30]

Finally, as an aside, we prove a corollary that will be useful later:

Corollary 3. *Suppose that an entrepreneur has a soft budget constraint. He is not captive if and only if*

$$(1+\rho)K_1 \leq \max\left\{p^2\frac{b}{\Delta}, (p-\Delta)\left(\frac{B}{\Delta} + p\frac{b}{\Delta}\right)\right\}. \tag{64}$$

Proof. To prove the result, we make use of the fact that increasing R_1 makes it easier to keep an entrepreneur captive (Proposition 2). Hence, an entrepreneur is captive if and only if a rival will not finance him given the maximum possible R_1 that satisfies

[30] The expression in the lemma can easily be verified case by case. To do so, it is useful to note that the conditions $\frac{b}{B} \leq \frac{1}{\Delta} - \frac{1}{p}$ and $\hat{Q} > \frac{p}{\Delta}\left(\frac{1}{\Delta} - \frac{1}{p}\right)$ imply that $\hat{Q}B > p\frac{b}{\Delta}$, and therefore that we are in Case (i) above.

his first-stage IC with the lowest continuation value consistent with an SBC, or with $u_1 = pb/\Delta$ and hence $R_1 = y - B/\Delta - pb/\Delta$. (Recall that Assumption 1 implies he works at the first stage, so the IC is satisfied.) To recover the condition in the corollary, we substitute this into the necessary and sufficient condition for the entrepreneur to be captive (inequality (10)).

□

A.5 Proof of Proposition 3

We divide the proof into a number of smaller results:

- Lemma 2 characterizes an entrepreneurs' project choice, given their access to finance.
- Lemma 3 characterizes non-bank finance for HAC projects (which turns out to be the only relevant case).
- Lemma 4 characterizes bank finance for HAC and LAC projects. There, we show that rival banks provide continuation financing to LAC entrepreneurs and, in Lemma 5, we show that rival non-banks do not.

Uniqueness follows from these results, because we show how each player best responds to any rationalizable action of others (i.e. we do not rely on players knowing the equilibrium behavior of others).

Lemma 2. *If an entrepreneur can finance either the HAC or the LAC project, he chooses the HAC project.*

Proof. An entrepreneur gets at least his first-stage agency rent, pB^A/Δ if he does the HAC project. He gets at most the sum of his first- and second-stage agency rents, which is at most $2pB^a/\Delta$, if he does the LAC project.[31] Since $B^A > 2B^a$ by Assumption 2, he always prefers the HAC project. □

Lemma 3. *Non-banks (i) impose HBCs on HAC entrepreneurs, (ii) keep them captive, and (iii) are willing to provide finance at the initial stage.*

Proof. We prove the three statements in turn.

- **Statement (i).** This follows from Proposition 1: inequality (9) is violated for $\alpha = A$ and $\rho = r$ by Assumption 4.

[31] To see this upper bound, observe that the entrepreneur's ex ante utility is lower if he shirks (and gets private benefits) than if he works (and gets agency rents which more than compensate for forgone private benefits). In this case, he gets $u_0 = p(y - R_1) + (1 - p)p(y - R_1 - R_2)$. From the first- and second-stage ICs (inequalities (8) and (6)), we have that $R_1 + R_2 \le y - B/\Delta$, so $u_1 \le pB/\Delta$, and $R_1 \le y - B/\Delta - u_1$. Substituting into the expression for u_0 gives the upper bound in the text.

- **Statement (ii).** This follows from Statement (i) and Corollary 1, given Assumption 5.
- **Statement (iii).** The non-bank's expected payoff is $p\,(y - B/\Delta)$, which is just the project's expected payoff py minus the entrepreneur's agency rent pB/Δ at the first stage (and nothing at the second stage, given his has an HBC and he is captive). This exceeds $(1+r)K_0$ (i.e. the non-bank's first-stage PC is satisfied) by Assumption 4.

□

Lemma 4. *Banks (i) have SBCs with both types of entrepreneurs, (ii) do not keep LAC-entrepreneurs captive, and (iii) are willing to provide finance at the initial stage to LAC- but not HAC-entrepreneurs.*

Proof. We prove the three statements in turn.

- **Statement (i).** This follows from Proposition 1: inequality (9) is satisfied for $\alpha \in \{A, a\}$ and $\rho = 0$ by Assumption 1 and Assumption 5.[32]
- **Statement (ii).** This follows from Corollary 3: given their SBCs, entrepreneurs are not captive to banks as long as inequality (64) is satisfied with $\rho = 0$, which it is by Assumption 5.
- **Statement (iii).** Here we compute the bank's expected payoff and show that it is positive with an LAC entrepreneur but negative with an HAC entrepreneur.

  Given the bank has an SBC and $\rho = 0$, the bank's payoff is

$$\begin{aligned}\text{bank's payoff } &= py + (1-p)py - K_0 - (1-p)K_1 - u_0 \\ &= (2-p)py - K_0 - (1-p)K_1 - u_0,\end{aligned} \tag{67}$$

  where $py + (1-p)py$ is the total surplus (given the entrepreneur works at both stages), $K_0 + pK_1$ is the total expected capital outlay, and u_0 is the entrepreneur's payoff. We can find upper and lower bounds on u_0. To do so, observe that from his first-stage IC (inequality (7)) the entrepreneur gets agency rent $p(B/\Delta + u_1)$ at stage one. Adding his continuation value gives $u_0 = pB/\Delta + u_1$. Now, from

[32] To see this, observe after that substituting $p^2 b > \Delta K_1$ from Assumption 5 and $B > b$, Assumption 1 implies that

$$y \geq \frac{1-\Delta}{(1-p)\Delta}\Big(p\frac{b}{\Delta} + K_1\Big). \tag{65}$$

Given $\Delta < p$, this implies that

$$y \geq \frac{1}{p}\left(p\frac{b}{\Delta} - K_1\right). \tag{66}$$

Rearranging gives the desired inequality.

Lemma 1, observe that his continuation value u_1 is at least pb/Δ and at most pB/Δ. Hence,

$$\frac{p}{\Delta}(B+b) \leq u_0 \leq \frac{2pB}{\Delta}. \tag{68}$$

Thus, substituting into the equation for the bank's payoff, we have that

$$(2-p)py - K_0 - (1-p)K_1 - \frac{2pB}{\Delta} \leq \text{ bank's payoff } \leq (2-p)py - K_0 - (1-p)K_1 - \frac{p(B+b)}{\Delta}. \tag{69}$$

By Assumption 3, the LHS is always positive for LAC projects (so a bank always finances them) and the RHS is always negative for HAC projects (so a bank never finances them).

□

Lemma 5. *Non-banks do not finance continuation of LAC entrepreneurs.*

Proof. The rival non-bank finances continuation if and only inequality (10) is satisfied with $\rho = r$.

A sufficient condition for this is that inequality (10) is satisfied with the lowest possible value of R_1, which is the value at which the first-stage IC binds with the largest possible value of u_1 ($\hat{Q} = 1$). This value is $u_1 = pB/\Delta$, which corresponds to the case in which the entrepreneur meets a rival that incentivizes him to work for sure.

From here, the binding first-stage IC (inequality (8)) gives $R_1 = y - (1+p)B/\Delta$. Substituting into the condition for captivity in inequality (10), we find that non-banks do not provide continuation financing if

$$(1+r)K_1 > \max\left\{p^2\frac{B}{\Delta}, (p-\Delta)(1+p)\frac{B}{\Delta}\right\}. \tag{70}$$

This is satisfied for the LAC entrepreneur by Assumption 5. □

A.6 Proof of Corollary 2

The result follows from Lemma 5. □

A.7 Proof of Proposition 4

We prove each statement in turn.

- **Statement 1.** We must show that for sufficiently low Q, all financiers prefer to become banks. Substituting $Q = 0$ into the continuation value u_1 in Lemma 1 ($u_1(0) = pb/\Delta$), we see that banks' payoff is greater than non-banks' (the expression in equation (22) with $Q = 0$ is greater than that in equation (20)) exactly when the condition in the proposition is satisfied (inequality (27)).

- **Statements 2.** This follows almost immediately from the analysis in the text given the expression in equation (26) for φ.
- **Statement 3.** Using the expression for φ in equation (26) and re-writing, we see that $\varphi < 1$ for all Q if and only if $u^* > u_1(0) = pb/\Delta$, where the last equality follows from the expression in Lemma 1. This is always satisfied given the definition of u^* in equation (24) and the condition in Proposition (27). □

A.8 Proof of Proposition 5

As above, an entrepreneur's budget constraint is soft with a bank. Thus, his payoff is

$$u_0 = p(y - R_1) + (1 - p)u_1 \tag{71}$$

$$= p\frac{B}{\Delta} + u_1, \tag{72}$$

having substituted for R_1 from the entrepreneur's IC (equation (8)).

Now, by the hypothesis of the proposition, he is captive with an HAC project, but not with an LAC one. Hence, $u_1\big|_{\text{HAC}} = p\frac{b^A}{\Delta}$ and $u_1\big|_{\text{LAC}}$ is a function of $\hat{Q}$ given by Lemma 1. Thus $u_0|_{\text{HAC}}$ is constant, whereas $u_0|_{\text{LAC}}$ is increasing in $\hat{Q}$. Hence, he chooses the LAC project if $\hat{Q}$ is sufficiently high. □

A.9 Proof of Proposition 6

The argument follows from taking the limit as $Q \to 1$ of the financiers' indifference condition. We have that $\hat{Q} = (1 - \varphi_c)Q \to 1 - \varphi_c$ and that

$$\lim_{Q\to 1} \frac{q_{\text{nb}}}{q_b} = \lim_{Q\to 1} \frac{\frac{1}{1+\frac{\varphi_c}{1-Q}}}{\frac{1}{1+\frac{1-\varphi_c}{1-Q}}} \tag{73}$$

$$= \frac{1 - \varphi_c}{\varphi_c}. \tag{74}$$

By Lemma 1 and the hypothesis that $\frac{b}{B} \leq \frac{1}{\Delta} - \frac{1}{p}$, $u_1(\hat{Q}) \to pB^a(1 - \varphi_c)/\Delta$. Thus, equation (33) can be written as

$$p\frac{B^a}{\Delta}(1 - \varphi_c) = u^* + \left(1 - \frac{1 - \varphi_c}{\varphi_c}\right)\left(\Sigma^A - p\frac{B^A}{\Delta}\right) \tag{75}$$

or, defining

$$\beta := \frac{\Sigma^A - p\frac{B^A}{\Delta}}{pB^a/\Delta}, \tag{76}$$

as

$$\varphi_c^2 - \left(1 - \frac{u^*}{pB^a/\Delta} - 2\beta\right)\varphi_c - \beta = 0. \tag{77}$$

The expression for φ_c in the proposition follows from solving the quadratic equation for φ_c and realizing that the smaller root is negative, and hence can be discarded.

To compare the above to the fraction of non-banks in the baseline model, observe from equation (24) that, for $u_1 \to pB^a(1-\varphi)/\Delta$, the limit of φ in the baseline model is

$$1 - \frac{u^*}{pB^a/\Delta} =: \varphi^\infty \tag{78}$$

Comparing this to the expression for φ_c (equation (34)) and manipulating reveals that the limiting φ_c exceeds φ^∞ if and only if $\varphi^\infty < 1/2$. □

A.10 Proof of Proposition 7

From equation (38), we have that $u_1(\hat{Q}) > u^*$ and, therefore, using $\hat{Q} = (1-\varphi_s)Q$ from Lemma 1, $\varphi_s < 1 - u^{-1}(u^*)/Q$. The RHS is the expression for $\varphi > 0$ in the baseline model (equation (26)). Hence, $\varphi_s < \varphi$.

It remains only to check that entrepreneurs still choose innovative projects when they meet non-banks, despite the risk that the projects are not viable. Given that entrepreneurs strictly prefer innovative projects in the baseline model by Lemma 2, this is the case as long as the probability of getting a viable project S^A/E^A is high enough, which it is given our assumption that S^A is not too small. □

B Assumption 1 Implies No Equilibrium Shirking

Here, we explain that financiers always offer contracts that incentivize work at both stages, i.e. that it is most profitable to offer repayments R_1 and R_2 that satisfy the entrepreneur's ICs (inequalities (5) and (7)). Note, however, that at the second stage the entrepreneur's outside option is to get finance from a rival, which might not offer a contract satisfying his IC.

HBC. If the entrepreneur faces an HBC, there is only one stage, and we need to show only that the financier's surplus is higher from working than from shirking, or,

$$p\left(y - \frac{B}{\Delta}\right) \geq (p-\Delta)y, \tag{79}$$

where the LHS is the financier's payoff if the entrepreneur's first-stage IC binds and the

RHS is its payoff if the feasibility constraint ($R_1 = y$) binds. Re-writing, this says that

$$\Delta^2 y \geq pB. \tag{80}$$

This is the (second part of the) condition in Assumption 1.

SBC. Now we compare the financier's payoff from incentivizing work or not at each stage. There are four possible outcomes, for each first and second stage action, respectively work-work, work-shirk, shirk-work, and shirk-shirk. We focus on the case of a bank, which turns out to be the only relevant case here (since, by Lemma 3, the non-bank imposes an HBC).

We already know from the proof of Proposition 1 that work-work $\succeq$ work-shirk. Here, we compute the financier's payoff from the other outcomes and show that work-work is necessarily preferred (no matter whether the entrepreneur has access to a rival).

- **Work-work.** Here, we use the superscript ww to indicate the repayments the financier offers such that the entrepreneur works at each stage, i.e. that satisfy the entrepreneur's ICs. In this case, a financier gets

$$\Pi_1^{\text{work,work}} = -K_0 + pR_1^{ww} + (1-p)\Big(- K_1 + p(R_1^{ww} + R_2^{ww})\Big). \tag{81}$$

  The terms can be understood as follows. The financier provides the initial capital K_0 and the entrepreneur works. Hence, he succeeds and repays with probability $\pi_1 = p$. He fails with probability $1-p$, in which case, his budget constraint is soft: the financier provides continuation capital K_1 and the entrepreneur works and, hence, succeeds and repays with probability $\pi_2 = p$.

- **Shirk-work.** Here, we use the superscript sw to indicate the repayments the financier offers such that the entrepreneur shirks at the first stage and works at the second, i.e. that satisfy the entrepreneur's resource constraint at the first stage ($R_1^{sw} = y$) and its IC at the second (inequality (5)). In this case, a financier gets

$$\Pi_1^{\text{shirk,work}} = -K_0 + (p-\Delta)y + (1-p+\Delta)\Big(- K_1 + p(R_1^{sw} + R_2^{sw})\Big). \tag{82}$$

  The terms can be understood as follows. The financier provides the initial capital K_0 and the entrepreneur shirks. Hence, he succeeds and repays with probability $\pi_1 = p-\Delta$. He fails with probability $1-p+\Delta$, in which case, his budget constraint is soft: the financier provides continuation capital K_1 and the entrepreneur works and, hence, succeeds and repays with probability $\pi_2 = p$.

- **Shirk-shirk.** In this case, a financier gets

$$\Pi_1^{\text{shirk,shirk}} = -K_0 + (p-\Delta)y + (1-p+\Delta)\Big(- K_1 + (p-\Delta)y\Big). \tag{83}$$

The terms can be understood as follows. The financier provides the initial capital K_0 and the entrepreneur shirks. Hence, he succeeds and repays with probability $\pi_1 = p - \Delta$. He fails with probability $1 - p + \Delta$, in which case, his budget constraint is soft: the financier provides continuation capital K_1 and the entrepreneur shirks and, hence, succeeds and repays with probability $\pi_2 = p - \Delta$.

We first point out that shirk-work $\succeq$ shirk-shirk: the first-stage payoff to the financier is the same, and we know from the proof of Proposition 1 that, given Assumption 1, financiers always prefer to induce working at the second stage.

It remains to show that work-work $\succeq$ shirk-work. To compare the expressions above, observe first that the total repayment at the second stage does not depend on what happened in the first stage, so $R_1^{ww} + R_2^{ww} = R_1^{sw} + R_2^{sw}$. Now, given the expressions above, we have

$$\Pi_1^{\text{work,work}} \geq \Pi_1^{\text{shirk,work}} \implies p\Big(y - R_1^{ww}\Big) < \Delta\Big(y + K_1 - p(R_1^{sw} + R_2^{sw})\Big). \tag{84}$$

This inequality is satisfied if it is satisfied for the lowest possible R_1^{ww} and the largest possible $R_1^{sw} + R_2^{sw}$.

- The lowest possible R_1^{ww} comes from, first, making the first-stage IC (inequality (8)) bind, so $R_1^{ww} = y - B/\Delta - u_1$, and, second, making u_1 as large as possible, so pB/Δ (which is an upper bound on the entrepreneur's second-stage payoff). Thus, we set $R_1^{ww} = y - B/\Delta - pB/\Delta$.
- The largest possible $R_1^{sw} + R_2^{sw}$ comes from making the second-stage IC (inequality (6)) bind, so $R_1^{sw} + R_2^{sw} = y - b/\Delta$.

Substituting R_1^{ww} and $R_1^{sw} + R_2^{sw}$ into inequality (84), we get a sufficient condition for there not to be shirking at the first stage. This is the (first part of the) condition in Assumption 1. □

References

Acharya, V., H. Almeida, F. Ippolito, and A. Perez (2014). Credit lines as monitored liquidity insurance: Theory and evidence. *Journal of Financial Economics 112*(3), 287 – 319.

Allen, F. and D. Gale (2004). Financial intermediaries and markets. *Econometrica 72*(4), 1023–1061.

Axelson, U., P. Strömberg, and M. Weisbach (2009). Why are buyouts levered? The financial structure of private equity firms. *Journal of Finance 64*, 1549–1582.

Barry, C. B. and V. T. Mihov (2015). Debt financing, venture capital, and the performance of initial public offerings. *Journal of Banking and Finance 58*, 144 – 165.

Begenau, J. and T. Landvoigt (2017). Financial regulation in a quantitative model of the modern banking system. Working paper, Harvard Business School.

Berger, A. and K. Schaek (2011). Small and medium-sized enterprises, bank relationship strength, and the use of venture capital. *Journal of Money, Credit and Banking 43*, 461–490.

Besanko, D. and G. Kanatas (1993). Credit market equilibrium with bank monitoring and moral hazard. *The Review of Financial Studies 6*(1), 213–232.

Biais, B. and A. Landier (2015). Endogenous agency problems and the dynamics of rents. Working paper, HEC Paris.

Blanchflower, D. and A. Oswald (1998). What makes an entrepreneur? *Journal of Labor Economics 16*(1), 26–60.

Bolton, P. and X. Freixas (2000). Equity, bonds, and bank debt: Capital structure and financial market equilibrium under asymmetric information. *Journal of Political Economy 108*(2), 324–351.

Bolton, P. and D. S. Scharfstein (1990, March). A theory of predation based on agency problems in financial contracting. *American Economic Review 80*(1), 93–106.

Bond, P. (2004). Bank and nonbank financial intermediation. *Journal of Finance 59*(6), 2489–2529.

Boot, A. and A. Thakor (1994). Moral hazard and secured lending in an infinitely repeated credit market game. *International Economic Review 35*(4), 899–920.

Boot, A. and A. Thakor (1997). Financial system architecture. *Review of Financial Studies 10*(3), 693–733.

Boot, A. W. A. and A. V. Thakor (2000). Can relationship banking survive competition? *Journal of Finance 55*(2), 679–713.

Botsch, M. and V. Vanasco (2019). Learning by lending. *Journal of Financial Intermediation 37*, 1 – 14.

Boualam, Y. (2018). Credit markets and relationship capital. Working paper, University of North Carolina.

Boyd, J. and G. De Nicolò (2005). The theory of bank risk-taking and competition revisited. *Journal of Finance 60*, 1329–1343.

Boyd, J. H. and M. Gertler (1994). Are banks dead? Or are the reports greatly exaggerated? *Quarterly Review* (Sum), 2–23.

Buchak, G., G. Matvos, T. Piskorski, and A. Seru (2018). Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics 130*(3), 453 – 483.

Cao, M. and S. Shi (2000). Screening, bidding, and the loan market tightness. Center for Financial Institutions Working Papers 00-09, Wharton School Center for Financial Institutions, University of Pennsylvania.

Carey, M., M. Post, and S. A. Sharpe (1998). Does corporate lending by banks and finance companies differ? Evidence on specialization in private debt contracting. *The Journal of Finance 53*(3), 845–878.

Cetorelli, N. (2004). Real effects of bank competition. *Journal of Money, Credit and Banking 36*(3), 543–558.

Chan, Y.-S., D. R. Siegel, and A. Thakor (1990). Learning, corporate control and performance requirements in venture capital contracts. *International Economic Review 31*(2), 365–81.

Chava, S., A. Oettl, A. Subramanian, and K. V. Subramanian (2013). Banking deregulation and innovation. *Journal of Financial Economics 109*(3), 759 – 774.

Chemmanur, T. J. and P. Fulghieri (1994). Investment bank reputation, information production, and financial intermediation. *Journal of Finance 49*(1), 57–79.

Chen, H., P. Gompers, A. Kovner, and J. Lerner (2010). Buy local? The geography of venture capital. *Journal of Urban Economics 67*(1), 90 – 102. Special Issue: Cities and Entrepreneurship.

Chernenko, S., I. Erel, and R. Prilmeier (2018). Nonbank lending. Working paper 2018-13, Dice Center.

Chrétien, E. and V. Lyonnet (2019). Traditional and shadow banks. Working paper, INSEE.

Cochrane, J. H. (2005). The risk and return of venture capital. *Journal of Financial Economics 75*(1), 3 – 52.

Cohan, P. (2013). Why does location matter? *MIT Entrepreneurship Review*, http://miter.mit.edu/why-does-location-matter/.

Cornaggia, J., Y. Mao, X. Tian, and B. Wolfe (2015). Does banking competition affect innovation? *Journal of Financial Economics 115*(1), 189–209.

Crémer, J. (1995). Arm's length relationships. *The Quarterly Journal of Economics 110*(2), 275–295.

Degryse, H. and P. V. Cayseele (2000). Relationship lending within a bank-based system: Evidence from European small business data. *Journal of Financial Intermediation 9*(1), 90–109.

Degryse, H. and S. Ongena (2005). Distance, lending relationships, and competition. *The Journal of Finance 60*(1), 231–266.

Dell'Ariccia, G. (2000). Learning by lending, competition, and screening incentives in the banking industry. Center for Financial Institutions Working Papers 10, Wharton School Center for Financial Institutions, University of Pennsylvania.

Dell'Ariccia, G. and R. Marquez (2004). Information and bank credit allocation. *Journal of Financial Economics 72*(1), 185 – 214.

Denis, D. J. and V. T. Mihov (2003). The choice among bank debt, non-bank private debt, and public debt: Evidence from new corporate borrowings. *Journal of Financial Economics 70*(1), 3 – 28.

Dewatripont, M. and E. Maskin (1995). Credit and efficiency in centralized and decentralized economies. *The Review of Economic Studies* (62).

Diamond, D. W. (1991). Monitoring and reputation: The choice between bank loans and directly placed debt. *Journal of Political Economy 99*(4), 689–721.

Diamond, D. W. and R. G. Rajan (2001). Banks and liquidity. *American Economic Review 91*(2), 422–425.

Diamond, W. (2019). Safety transformation and the structure of the financial system. Working paper, Wharton School.

Donaldson, J. R. and G. Piacentino (2019). Money runs. Working paper, Washington University in St. Louis.

Donaldson, J. R., G. Piacentino, and A. Thakor (2018). Warehouse banking. *Journal of Financial Economics 129*(2), 250 – 267.

Donaldson, J. R., G. Piacentino, and A. Thakor (2019). Household debt overhang and unemployment. *The Journal of Finance 0*(0).

Duffie, D., N. Gârleanu, and L. H. Pedersen (2005). Over-the-counter markets. *Econometrica 73*(6), 1815–1847.

Gersbach, H. and H. Uhlig (2007). On the coexistence of banks and markets. *Scandinavian Journal of Economics 109*(2), 225–243.

Gompers, P. and J. Lerner (2001, June). The venture capital revolution. *Journal of Economic Perspectives 15*(2), 145–168.

Gompers, P. A. (1995). Optimal investment, monitoring, and the staging of venture capital. *The Journal of Finance 50*(5), 1461–1489.

Guler, I. (2018). Pulling the plug: The capability to terminate unsuccessful projects and firm performance. *Strategy Science 3*(3), 481–497.

Guzman, M. G. (2000). Bank structure, capital accumulation and growth: A simple macroeconomic model. *Economic Theory 16*(2), 421–455.

Hall, R. and S. Woodward (2007, April). The incentives to start new companies: Evidence from venture capital. Working Paper 13056, National Bureau of Economic Research.

Hall, R. and S. Woodward (2010). The burden of the nondiversifiable risk of entrepreneurship. *American Economic Review* (100), 1163–1194.

Hanson, S. G., A. Shleifer, J. C. Stein, and R. W. Vishny (2015). Banks as patient fixed-income investors. *Journal of Financial Economics 117*(3), 449–469.

Hellmann, T. and M. Puri (2000, 06). The interaction between product market and financing strategy: The role of venture capital. *The Review of Financial Studies 13*(4), 959–984.

Hellmann, T. F., K. C. Murdock, and J. E. Stiglitz (2000). Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough? *American Economic Review 90*(1), 147–165.

Herkenhoff, K. F. (2019). The impact of consumer credit access on unemployment. NBER working paper no. 22846.

Holmström, B. and J. Tirole (1997). Financial intermediation, loanable funds, and the real sector. *The Quarterly Journal of Economics 112*(3), 663–91.

Hombert, J. and A. Matray (2016). The real effects of lending relationships on innovative firms and inventor mobility. *The Review of Financial Studies 30*(7), 2413–2445.

Hoshi, T., A. Kashyap, and D. Scharfstein (1993). The choice between public and private debt: An analysis of post-deregulation corporate financing in Japan. NBER Working Paper 4421, National Bureau of Economic Research, Inc.

Hsu, P.-H., X. Tian, and Y. Xu (2014). Financial development and innovation: Cross-country evidence. *Journal of Financial Economics 112*(1), 116 – 135.

Inderst, R. and H. M. Mueller (2004). The effect of capital market characteristics on the value of start-up firms. *Journal of Financial Economics 72*(2), 319–356.

Inderst, R., H. M. Mueller, and F. Münnich (2006, 10). Financing a portfolio of projects. *The Review of Financial Studies 20*(4), 1289–1325.

International Monetary Fund (2016). *Global Financial Stability Report—Fostering Stability in a Low-Growth, Low-Rate Era*. Washington.

Irani, R., R. Iyer, R. Meisenzahl, and J.-L. Peydró (2020). The rise of shadow banking: Evidence from capital regulation.

Jensen, M. C. (1989). Eclipse of the public corporation. *Harvard Business Review 67*, 61–74.

Jensen, M. C. and W. H. Meckling (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics 3*(4), 305 – 360.

Jovanovic, B. and B. Szentes (2013). On the market for venture capital. *Journal of Political Economy 121*(3), 493 – 527.

Kashyap, A. K., R. Rajan, and J. C. Stein (2002). Banks as liquidity providers: An explanation for the coexistence of lending and deposit-taking. *The Journal of Finance 57*(1), 33–73.

Kaviani, M. and H. Maleki (2018). The conservative corporation: Banking competition and decline in firms' risk-taking. Working paper, Temple University.

Keeley, M. C. (1990, December). Deposit insurance, risk, and market power in banking. *American Economic Review 80*(5), 1183–1200.

Kerr, W. R., J. Lerner, and A. Schoar (2011, 10). The consequences of entrepreneurial finance: Evidence from angel financings. *The Review of Financial Studies 27*(1), 20–55.

Kerr, W. R., R. Nanda, and M. Rhodes-Kropf (2014, Summer). Entrepreneurship as experimentation. *Journal of Economic Perspectives 28*(3), 25–48.

Khanna, N. and R. Mathews (2017). The limits of reputation. Working paper, Michigan State University.

Kornai, J. (1979). Resource-constrained versus demand-constrained systems. *Econometrica* (47), 801–819.

Kornai, J. (1980). *The Economics of Shortage*. New York: North-Holland.

Korteweg, A. and S. Nagel (2016). Risk-adjusting the returns to venture capital. *The Journal of Finance 71*(3), 1437–1470.

Korteweg, A. and M. Sorensen (2010). Risk and return characteristics of venture capital-backed entrepreneurial companies. *The Review of Financial Studies 23*(10), 3738–3772.

Kortum, S. and J. Lerner (2000). Assessing the impact of venture capital to innovation. *RAND Journal of Economics 31*(4), 674–692.

Lerner, J., A. Schoar, S. Sokolinski, and K. Wilson (2015, December). The globalization of angel investments: Evidence across countries. Working Paper 21808, National Bureau of Economic Research.

Lerner, J., M. Sorensen, and P. Strömberg (2011). Private equity and long-run investment: The case of innovation. *The Journal of Finance 66*(2), 445–477.

Lim, J., B. A. Minton, and M. S. Weisbach (2014). Syndicated loan spreads and the composition of the syndicate. *Journal of Financial Economics 111*(1), 45–69.

Manso, G. (2011). Motivating innovation. *The Journal of Finance 66*(5), 1823–1860.

Mao, Y. and J. J. Wang (2018). Access to finance and technological innovation: Evidence from antebellum america. Working paper, Arizona State University.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science 2*(1), 71–87.

Marquez, R. (2002). Competition, adverse selection, and information dispersion in the banking industry. *Review of Financial Studies 15*(3), 901–926.

Martinez-Miera, D. and R. Repullo (2010). Does competition reduce the risk of bank failure? *Review of Financial Studies 23*(10), 3638–3664.

Martinez-Miera, D. and R. Repullo (2017). Search for yield. *Econometrica 85*(2), 351–378.

Matutes, C. and X. Vives (2000). Imperfect competition, risk taking, and regulation in banking. *European Economic Review 44*(1), 1–34.

Merton, R. C. and R. T. Thakor (2018). Customers and investors: A framework for understanding the evolution of financial institutions. *Journal of Financial Intermediation*.

Metrick, A. and A. Yasuda (2010). The economics of private equity funds. *Review of Financial Studies 23*(6), 2303–2341.

Mortensen, D. T. and C. A. Pissarides (1994). Job creation and job destruction in the theory of unemployment. *The Review of Economic Studies 61*(3), 397–415.

Nagel, S. (2016, 07). The liquidity premium of near-money assets. *The Quarterly Journal of Economics 131*(4), 1927–1971.

Nakamura, L. I. and K. Roszbach (2018). Credit ratings, private information, and bank monitoring ability. *Journal of Financial Intermediation 36*(C), 58–73.

Nanda, R. and M. Rhodes-Kropf (2013). Innovation and the financial guillotine. NBER Working Paper 19379, National Bureau of Economic Research Inc.

Neuhann, D. and F. Saidi (2016). Bank deregulation and the rise of institutional lending. Working paper, UT–Austin.

Nguyen, H.-L. Q. (2019, January). Are credit markets still local? Evidence from bank branch closings. *American Economic Journal: Applied Economics 11*(1), 1–32.

Payne, J. (2018). The disruption of long term bank credit. Working paper, New York University.

Petersen, M. A. and R. G. Rajan (1995). The effect of credit market competition on lending relationships. *The Quarterly Journal of Economics 110*(2), 407–43.

Piacentino, G. (2019). Venture capital and capital allocation. *The Journal of Finance (forthcoming) 0*(0).

Puri, M. and R. Zarutskie (2012). On the life cycle dynamics of venture-capital- and non-venture-capital-financed firms. *The Journal of Finance 67*(6), 2247–2293.

Rajan, R. G. (1992). Insiders and outsiders: The choice between informed and arm's-length debt. *Journal of Finance 47*(4), 1367–400.

Remolona, E. and K. Wulfekuhler (1992). Finance companies, bank competition, and niche markets. *Federal Reserve Bank of New York Quarterly Review* (2).

Repullo, R. (2004). Capital requirements, market power, and risk-taking in banking. *Journal of Financial Intermediation 13*(2), 156–182.

Repullo, R. and J. Suarez (2000). Entrepreneurial moral hazard and bank monitoring: A model of the credit channel. *European Economic Review 44*, 1931–1950.

Rin, M. D., T. Hellmann, and M. Puri (2013). Chapter 8 - A survey of venture capital research. Volume 2 of *Handbook of the Economics of Finance*, pp. 573 – 648. Elsevier.

Sahlman, W. A. (1990). The structure and governance of venture-capital organizations. *Journal of Financial Economics 27*(2), 473–521.

Samila, S. and O. Sorenson (2011). Venture capital, entrepreneurship, and economic growth. *The Review of Economics and Statistics 93*(1), 338–349.

Sharpe, S. A. (1990). Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships. *Journal of Finance 45*(4), 1069–87.

Song, F. and A. V. Thakor (2010). Financial system architecture and the co-evolution of banks and capital markets. *The Economic Journal 120*(547), 1021–1055.

Sorenson, O. and T. E. Stuart (2001). Syndication networks and the spatial distribution of venture capital investments. *American Journal of Sociology 106*(6), 1546–1588.

Startz, R. (1979). Implicit interest on demand deposits. *Journal of Monetary Economics 5*(4), 515 – 534.

Stevens, M. (2007). New microfoundations for the aggregate matching function. *International Economic Review 48*(3), 847–868.

Stiroh, K. and P. E. Strahan (2003). Competitive dynamics of deregulation: Evidence from U.S. banking. *Journal of Money, Credit and Banking 35*(5), 801–28.

Ueda, M. (2004). Banks versus venture capital: Project evaluation, screening, and expropriation. *Journal of Finance 59*(2), 601–621.

von Thadden, E.-L. (1999). Liquidity creation through banks and markets: Multiple insurance and limited market access. *European Economic Review 43*(4-6), 991–1006.

Wagner, W. (2009). Loan market competition and bank risk-taking. *Journal of Financial Services Research 37*.

Wasmer, E. and P. Weil (2004). The macroeconomics of labor and credit market imperfections. *American Economic Review 94*(4), 944–963.

Wessel, M. (2013). Don't build your startup outside of Silicon Valley. https://hbr.org/2013/10/dont-build-your-startup-outside-of-silicon-valley.

Zwiebel, J. (1995). Corporate conservatism and relative compensation. *Journal of Political Economy 103*(1), 1–25.