Journal of Banking and Finance 87 (2018) 318–332

Contents lists available at ScienceDirect

Journal of Banking and Finance

journal homepage: www.elsevier.com/locate/jbf

Cash flows and credit cycles

Nicolás Figueroa[a], Oksana Leukhina[b,*]

[a] Pontificia Universidad Católica de Chile, Chile
[b] Department of Economics, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4M4, Canada

ARTICLE INFO

Article history:
Received 11 October 2015
Accepted 21 October 2017

JEL classification:
E32
E44
G20

Keywords:
Macroeconomic instability
Financial instability
Lending standards
Adverse selection

ABSTRACT

Aggregate productivity falls in recessions and rises in expansions. Several empirical studies suggest that the systematic behavior of lending standards, with laxer (tighter) standards applied during expansions (recessions), is responsible for reverting trends in aggregate productivity. We build a dynamic model that rationalizes these findings. Adverse selection in credit markets emerges as a potential source of macroeconomic instability. The key idea modeled is that in order to effectively signal their type to financiers, productive entrepreneurs must suffer a cost. The effective cost of signaling rises with higher cash flow brought about by stronger economic fundamentals, because higher cash flow makes it easier for the unproductive type to mimic the productive type. Competition among the financiers then results in suboptimally lax lending standards. Low productivity entrepreneurs obtain financing, the producer composition effect inducing a recession. This, in turn, creates conditions – weak economic fundamentals and low cash flow – conducive to the emergence of tighter lending terms, the strong composition effect leading to an economic recovery.

1. Introduction

This paper investigates the dynamic interaction between financial markets and macroeconomic fluctuations. Several empirical studies document that lending standards, i.e. contractual arrangements used to screen borrowers, are eased in expansions and tightened in recessions, such systematic behavior of lending terms influencing aggregate productivity dynamics (e.g. Asea and Blomberg, 1998; Berger and Udell, 2004; Lown and Morgan, 2006). These studies suggest that laxer standards during economic booms allow for the unproductive firms to be funded, reducing aggregate productivity through the producer composition effect. On the contrary, tight lending standards during economic downturns tend to exclude bad projects, thus sowing the seeds of an economic recovery. These studies also support the popular view that credit markets create economic instability through the producer composition effect (e.g. Kindleberger, 1996).

In a recent work, Myerson (2012) argues there is a pressing need for applying the insights from microeconomic theory of credit markets to macroeconomic models of business cycles. There is undoubtedly a need for a deeper understanding of the forces underlying macroeconomic instability, the financial sector being one potential culprit especially in light of the recent financial crisis. To this end, we propose a simple dynamic model with endogenous lending standards and aggregate productivity, which allows to study their interaction.

The model dynamics is consistent with the empirical behavior of lending standards, default rates, entrants' quality, investment and cash flows over the business cycle. In sharp contrast to related theoretical literature – where instability is driven by some changes taking place at producer level (intensive margin) – instability in our model is driven by *producer composition dynamics* (extensive margin)[1] (Section 2). Our focus on the extensive margin is motivated by the empirical behavior of lending standards.

Our model features private types of entrepreneurs and a competitive credit market. The key idea, novel to this context, is that in order to effectively signal their type to the financiers, productive entrepreneurs must suffer a cost. Effective signals are costly. Indeed, if costless, signals would be easily mimicked by the unproductive types. The main insight that emerges from the model is that the signal cost is related to cash flow, and therefore, economic fundamentals. In times of high cash flow, it is easier for the unproductive type to mimic the productive types, and therefore the signal cost is greater.

To clarify the intuition, suppose there are two types of entrepreneurs, good (G) and bad (B). Type B is unproductive, has little

* Corresponding author.
E-mail addresses: nicolasf@uc.cl (N. Figueroa), oksana.m.leukhina@gmail.com (O. Leukhina).

[1] Matsuyama (2013) is one exception.

https://doi.org/10.1016/j.jbankfin.2017.10.013

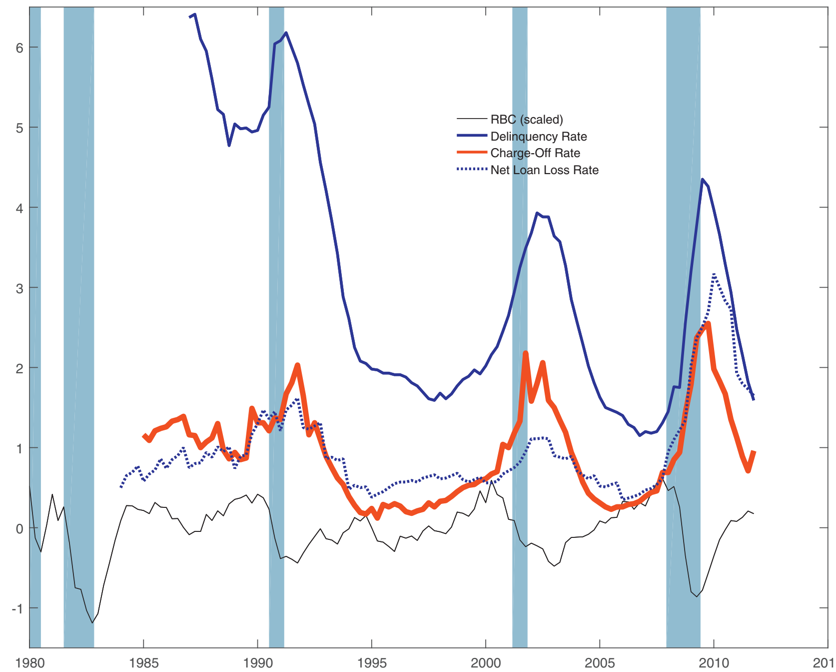

Fig. 1. Real business cycles, delinquency and charge-off rates on C&I loans *Notes*: the shaded areas are NBER-marked recessions. Real business cycles (RBC) are calculated by detrending the log of quarterly GDP data via the Hodrick–Prescott filter ($\lambda = 1600$).[35] Charge-off, delinquency and net loan loss rates are calculated for all business loans for all U.S. commercial banks. Charge-off rate on business loans is the value of business loans removed from the books and charged against loss reserves, net of recoveries, measured as a percentage of average total loans, and annualized. Delinquency rate is the value of business loans that are past due at least thirty days, measured as a percentage of average total loans. Net Loan Loss rate is the value of loan losses, net of recoveries, measured as a percentage of average total loans.[36]

to lose in case of default, and therefore always chooses to default. Contracts that exclude type *B* (separating contracts) necessarily restrict investment levels for type *G* – this is the cost of effective signaling. This cost endogenously increases with economic conditions: as cash flow rises, tighter investment restrictions are needed in order to exclude type *B*. The particularly high signaling cost makes separating contracts unattractive to type *G* in times of high cash flow. Competition for the productive types in the financial sector therefore results in pooling contracts, type *B* entrepreneurs entering as a side effect.

Lending standards are suboptimally lax (relative to the standards imposed by the planner with the same informational constraints) when the economic conditions are strong, which allows for greater investment at producer level, but also implies a misallocation of funds. Low productivity entrepreneurs are financed along with high productivity types, the producer composition effect sending the economy into a recession. This, in turn, creates conditions – weak economic fundamentals and low cash flow – conducive to the emergence of tighter lending terms, the strong composition effect leading to an economic recovery. Endogenous cycles may emerge, highlighting asymmetry of information in credit markets as a potential source of instability.

Rajan (1994) provides evidence from financial press and bankers' opinions showing that projects of negative present value are increasingly funded in economic expansions.[2] This view is also reflected in business loan delinquency rates beginning to rise while the economy is still expanding (Fig. 1). Policy makers have also expressed concerns regarding lax lending terms in times of strong economic fundamentals, pointing out that as much as 80% of new loan applications get approved without a formal projection of the borrowers' future performance (e.g. Supervisory letter SR 98-18 of the Board of Governors). Consistent with our model's insight, the letter quotes "intense competition to attract customers" combined with the presence of adverse selection as the reason behind the lending terms easing. Our explanation complements the explanation in Rajan (1994), based on banks' rational manipulation of current earnings. Note that information asymmetry is crucial for understanding the suboptimally lax standards and financing of ex-ante bad projects.

Ideally, if loan-level data were publicly accessible, we would explore the importance of cohort effects in accounting for the delinquency rate dynamics. The aforementioned papers some of which work with loan-level data would argue that delinquency rates increase during expansions at least partly because of the most recent borrowers, i.e. selection effects. After all, strong economic conditions should benefit the incumbent borrowers. To drive our point further though, we obtained data on Moody's rated corporate bonds and loans which exhibit similar delinquency dynamics over the cycle. Specifically, we use cumulative issuer-weighted default rates by year and by annual cohort (1970–2008)[3] to estimate cohort effects on default probability. To do so, we regress the year-specific default rate on the cohort age, age squared, economic conditions[4] and the cohort effect. Fig. 2 plots the obtained cohort effects together with the business cycle at the time of debt issuance. The data suggest that debt issued near the end of economic expansions is most likely to be defaulted upon.

In Section 4, we discuss a simple extension to three types of entrepreneurs, which more clearly marks the periods of economic expansions and recessions thereby helping illustrate the empirical relevance of the mechanism. The dynamics of cash flows, investment, credit lines and default rates all conform to their empirically observed behavior along the cycle. Expansions last longer than recessions, as in the data. In our model, the relatively high entrant quality in troughs is a result of tight financing terms. Consis-

[2] Financing of ex-ante bad projects is probably best exemplified with the dot-com mania of the late nineties and the recent subprime mortgage lending.

[3] The data is retrieved from Moody's Special Comment, Global Credit Policy – Corporate Default and Recovery Rates, 1920–2008, Februrary 2009.

[4] We use the contemporaneous change in the real business cycle to proxy for the economic conditions.

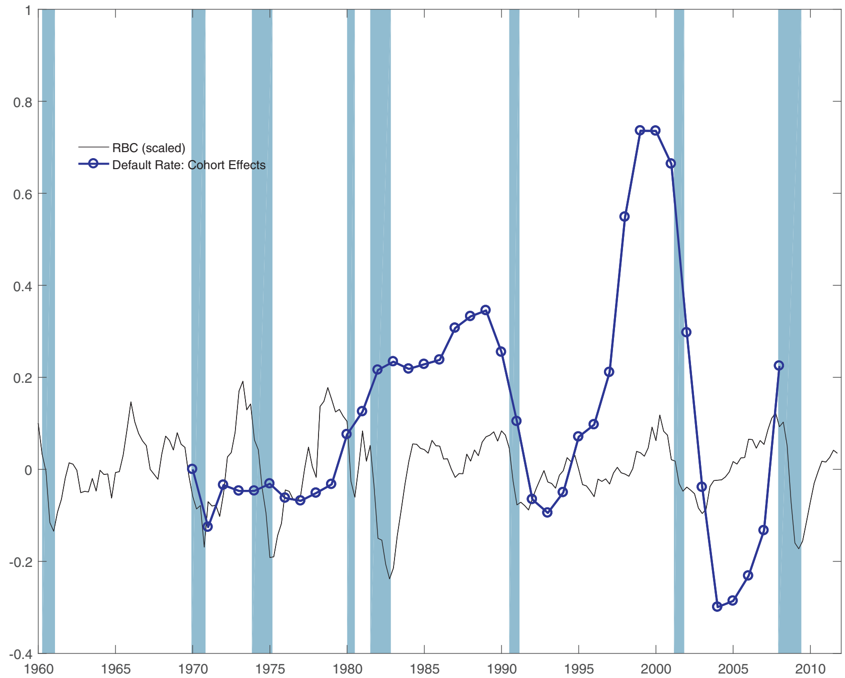

Fig. 2. Real business cycles and cohort effects on default of Moody's rated securities. *Notes*: Moody's rated corporate bonds and loan issuers. Specifically, we use cumulative issuer-weighted default rates by year and by annual cohort (1970–2008)[37] to document cohort effects on default probability. To do so, we regress the year-specific default rate on the cohort age, age squared, economic conditions[38] and the cohort effect. The figure plots the obtained cohort effects together with the business cycle at the time of debt issuance. The data suggest that debt issued near the end of economic expansions is most likely to be defaulted upon.

tent with our model's implications, Lee and Mukoyama (2012) document that manufacturing plants that enter under weak economic conditions are of significantly higher quality.[5] Conversely, the worst types enter production at the very peak. This producer composition dynamics along the cycle is what allows the model to provide insight into the cyclical behavior of default rates, observed in the data.

Our assumption of adverse selection best captures the situation in financial markets (both debt and equity finance) for informationally opaque firms where collateralizable wealth is not readily available. These markets serve relatively recent entrants along with mature businesses that lack audited financial statements.[6] Although it is impossible to quantify the precise contribution of these firms to aggregate productivity variation over the business cycles, it is clear that their contribution is important. One piece of evidence is that startups account for a large part of the cyclical component of job creation, a statistics measured with less ambiguity than productivity. Lee and Mukoyama (2012) document that, in manufacturing, job creation by startups in booms exceeds the net job creation of preexisting firms. More generally, small business accounts for 60–80% of the aggregate job creation/destruction.[7]

Our model is consistent with Hyman Minsky's idea that financial markets "cause" economic instability (e.g. Minsky, 1992). As in his financial instability hypothesis, recently popularized by media coverage, periods of economic expansions result in lower lending standards applied by profit-seeking banks, which consequently leads to the entry of low quality borrowers, upsurge in default rates and onset of economic recessions. However, Minsky's theory applies to financial markets where collateral is available and heavily relies on speculative beliefs regarding future collateral values. Our theory is complementary because it applies to financial markets where collateral is unavailable and highlights forces that do not rely on speculative expectations – all agents are perfectly rational.

The rest of the paper is organized as follows. Section 2 overviews related literature. In Section 3, we introduce the general model, derive static equilibrium contracts for given prices and define the dynamic general equilibrium. In Section 4, we study a fully dynamic economy with externalities in the production sector and analyze endogenous lending terms and productivity cycles. Production externality is assumed for the purpose of shutting down cash flow movements due to the price of capital. This greatly simplifies our analysis without compromising much insight.[8] In Section 5, we show that the results are robust to endogenizing the amount of loanable funds. We conclude in Section 6.

2. Related literature

Literature that brings together credit market frictions and economic activity predominantly focuses on the role of financial markets in amplifying exogenous productivity shocks.[9] While the amplification mechanism is also present in our model, our emphasis is on the existence of the opposite effect working through the composition of producers. A prominent amplification mechanism is due to Bernanke and Gertler (1989), where the borrowers' balance sheets amplify exogenous external shocks in a model of

[5] This can be thought of as a variant of the cleansing effect of recessions transpiring through the entry margin. The traditional view of cleansing roles of recessions envisioned recessions as opportunities to engage in restructuring and has been challenged by Caballero and Hammour (2005).

[6] Similar systematic behavior of equity financing terms is documented in venture capitalist markets (Gompers et al., 2008), and we show later that our model can be also applied to these markets.

[7] Small business is formally defined as firms with fewer than five hundred employees. The data are retrieved from the U.S. Small Business Administration, Office of Advocacy.

[8] The general model with no externalities was analyzed in an earlier version of this paper. See supplemental notes, Figueroa and Leukhina (2015) for the results.

[9] The early theoretical models on informational frictions in credit markets are Stiglitz and Weiss (1981), Bester (1985) and De Meza and Webb (1987).

costly state verification.[10] Economic upturns improve the borrowers' net worth, which lowers agency costs of financing investment, increases investment and hence amplifies upturns. The increase in output comes through both the intensive margin (output per producer) and extensive margin (more producers). A similar amplification mechanism is also present in our model at the intensive margin: higher cash flow leads to less screening and hence greater output per producer. In Kiyotaki and Moore (1997), enforcement issues imply that debt is collateralized. A temporary shock that increases a credit-constrained firm's net worth enables more borrowing, investment rises propagating the effect of the shock.[11] In Rampini (2004), entrepreneurial activity, which consists of engaging in risky but productive, in expected terms, projects increases at peaks. Other studies in this strand include Williamson (1987), Greenwald and Stiglitz (1993), and Bernanke et al. (1999).[12]

Our paper is closer to the strand of literature that highlights the role of credit market frictions in *reverting* aggregate productivity trends. In contrast to our paper, this literature almost exclusively focuses on the intensive margin, that is, productivity at producer level responds to the economic fundamentals in a way that generates economic instability. Most papers emphasize the role of moral hazard. Suarez and Sussman (1997), for example, extend Stiglitz and Weiss (1981) to include three overlapping generations. During booms, old entrepreneurs sell high quantities and, as a consequence, prices are low, implying low revenues for the young entrepreneurs and therefore a greater need for external financing. This generates excessive risk-taking at producer level, booms leading to busts. In Reichlin and Siconolfi (2004), entrepreneurs have a choice between safe and risky projects, the latter yield less in expected terms due to higher setup costs. When loanable funds are abundant, credit markets induce risky production, higher setup costs leading to a recession. In Favara (2012), monitoring is relaxed when entrepreneurs' net worth is high, because financiers' exposure to risk is low. This prompts entrepreneurs to engage in projects that yield higher private benefits but less profit, sending the economy into a recession.[13]

Also belonging to this strand, Azariadis and Smith (1998) and Martin (2008) highlight the role of adverse selection in generating instability via its effect on producer-level investment, rather than via its effect on producer composition as in this paper. The former presents financial regime switches driven by changes in savers' expectations of the future interest rate. In the latter work, instability arises due to the presence of an interesting tradeoff between loan size and collateral as screening tools. Separating contracts with stricter lending terms arise at the top of the cycle, implying low investment. The low investment then causes the downturn in the economy. The opposite is true in our model: lending standards are lax and investment is high at the top of the cycle, consistent with empirical evidence. It is the entry of low productivity entrepreneurs at the top that drives the economy down.

Our paper complements this strand by emphasizing the role of adverse selection in generating economic instability via its effect on producer composition – the extensive margin. Our focus on producer composition is motivated by the systematic behavior or lending terms observed in the data, which is bound to affect producer composition. The importance of compositional effects in generating economic instability is also emphasized in Matsuyama (2013). In that paper, projects differ in the externality they generate for the net worth of future entrepreneurs. Projects which do not generate any positive externality are subject to the borrowing constraint. Therefore, relatively more (less) credit is allocated to these projects when the net worth is high (low). In the present study, compositional effects are driven by qualitative changes in the nature of financial contracts, as determined by the problem of informational asymmetry.

Although our goal is to study the interaction of lending standards and economic conditions, we also contribute to the strand of literature that focuses on explaining the strong dependence of lending terms on economic conditions. The explanation in Berger and Udell (2004) invokes changing ability of loan officers. In Dell'ariccia and Marquez (2006), the cost of screening rises in expansions due to an influx of unknown borrowers. In Ruckes (2004), the pool of applicants at the top of the cycle is too good to warrant costly screening. Figueroa and Leukhina (2015) emphasize the role of production complementarity between economic conditions and the desired loan size, which makes screening too expensive in times of strong economic fundamentals.

3. The model

3.1. Environment

Consider a model economy where time is discrete and indexed by $t = 0, 1, 2, \ldots$. It is populated with overlapping generations of entrepreneurs who live for two periods and there exist two types of goods: consumption and capital. Generations represent the length of a financial contract. Only when young, entrepreneurs are endowed with a unit of time and ability to implement projects that produce capital. They enjoy utility from consuming when young and old, according to homothetic preferences represented by $u(c^y, c^o)$.[14] A savings technology is available to them at the risk free rate R_f.

The consumption good is produced by a continuum (of measure 1) of infinitely lived competitive firms. Further assuming a constant returns to scale technology, $F(K_t, L_t, t) = A_t K_t^{\beta} L_t^{1-\beta}$, allows us to restrict attention to an aggregate firm. We assume that capital depreciates upon use and that capital used in period t production must be purchased from young entrepreneurs in period $t-1$, at the price ρ_t. The aggregate firm can borrow at the risk free rate R_f.[15]

Capital goods are produced by young entrepreneurs of unobservable type $i \in \{G, B\}$ that differ in productivity. Each generation consists of measure μ of type G and measure $1-\mu$ of type B. Each entrepreneur can implement a project within a single period, but in two stages. Projects do not require time, and therefore all entrepreneurs will inelastically supply their unit of time to the consumption good sector, earning w_t. The production technology is linear, transforming investment (in terms of the consumption good) into capital at the rate f_i in the first stage and g_i in the second stage. Both stages are subject to a 1 unit minimum investment. We assume that type G is always more productive than type B and that projects are relatively unproductive in stage 1. These assumptions capture the essence of entrepreneurial production, i.e. that there are up-front costs that need to be covered prior to seeing positive returns to investment. More importantly, the presence

[10] Carlstrom and Fuerst (1997) study the mechanism in Bernanke and Gertler (1989) quantitatively.

[11] See Cordoba and Ripoll (2004) for quantitative analysis.

[12] Several studies study the impact of shocks to credit markets on economic activity (See Gertler and Kiyotaki, 2010). Khan and Thomas (2017) study the effects of a temporary tightening of lending standards. Jermann and Quadrini (2012) view financial sector shocks as a source of business cycles.

[13] In Myerson (2012) and Boissay et al. (2013), macroeconomic instability is driven by moral hazard in financial intermediation.

[14] Consumption takes place at the end of each period. Homotheticity is needed to facilitate the model extension in Section 6.

[15] Because the optimal input K_t must be determined in period $t-1$, at the time of its purchase, the firm must then form beliefs about the next period wage w_t. Given the price ρ_t and beliefs w_t, the optimal inputs K_t, L_t solve $\max_{K_t, L_t} A_t K_t^{\beta} L_t^{1-\beta} - R_f \rho_t K_t - w_t L_t$.

of the two production stages is what will allow the banks to differentiate between the two types based on their initial cash flows. Formally,

Assumption 1. Type G is more productive: $f_B < f_G$, $g_B < g_G$.

Assumption 2. Second stage production is relatively more productive: $\rho_t f_i < 1 < \rho_t g_i$.[16]

There is a competitive banking sector that loans investment funds to young entrepreneurs and has access to a risk free savings technology at rate R_f. Each period, banks are endowed with 2μ units of loanable funds, exactly the amount of the minimum investment needed to implement all type G projects. This particular amount is assumed for analytical simplicity. We relax this assumption by endogenizing the supply of loanable funds in Section 5. A risk-free savings technology is available to the banks at rate R_f.

Loan contracts are signed in the beginning of entrepreneurs' lives. We assume that whenever financing is provided, banks can ensure that entrepreneurs engage in production, i.e. invest the minimum required amount. This can be justified by the availability of a monitoring technology. The lack of collateralizable wealth and business idea screening technology severely restricts the ability of banks to screen entrepreneurs. This leads us to consider loan contract offers of the form $(\delta_{t,i}, R_{t,i})$. If a young entrepreneur i enters into a contract characterized by $(\delta_{t,i}, R_{t,i})$, he receives 1 unit of funds to be used in the first stage of production. He receives another unit of funds in the beginning of the second stage, conditional on paying δ_i towards the loan balance, and owes the bank the remaining balance $2R_{t,i} - \delta_{t,i}$ at the end of the period. Ability to pay $\delta_{t,i}$ is informative of one's cash flow, and therefore allows for effective screening.

It is important to note that we lose no generality by considering contracts that pay 1 unit at each stage. A more general contract would specify the amounts loaned at each stage, and we show in the first proof of the appendix that allowing for more general contracts will lead to equilibrium contract offers that pay 1 unit at each stage.[17]

The early payment $\delta_{t,i}$ captures the key idea of this paper – that a *costly* signaling device is needed to gather "hard evidence" about the entrepreneur's type.[18] The early payment $\delta_{t,i}$ constrains overall investment by limiting the amount of cash flows that can be reinvested in second stage production.

What kind of contract menus $\{(\delta_{t,i}, R_{t,i})\}_i$ are offered in equilibrium is discussed below. We will interpret $\{\delta_{t,i}\}_i$ observed in equilibrium as the stylized version of lending standards.

In addition, we assume that type B agents have the ability to abscond with the proceedings of their production at the end of the period, and therefore never repay their loans. Our assumption that default can take place only at the end of the second stage captures the idea that it is more difficult to hide income during initial stages of production.[19] Type G agents lack the ability to abscond with money and repay in full.[20]

The financial structure of capital producing firms is irrelevant in our environment, because project returns are deterministic.[21] Whether the financing is done through debt, equity, or a mix of both, the results go through. Equity contracts would be characterized with dividend payouts at the end of each stage. Discrimination of types would be done via the first dividend payout. Therefore, our setup is applicable to equity markets with adverse selection such as, for example, venture capital markets.

As a final remark regarding the model setup, we designed a simple theory which allocates an important role to the interaction of cash flow and financial markets. With cash flow of time t capital producers (as determined by w_t or ρ_{t+1}) evolving endogenously through time, the cost of effective signaling will also evolve endogenously through time, determining the level of screening employed in equilibrium at a particular point in time. Although we chose to model cash flow as arising from the wage and capital income, a number of other modeling choices would suffice (e.g. income from land holdings).

3.2. Entrepreneurs' optimization

We now discuss the entrepreneurs' choices and the static financial market equilibrium. A young entrepreneur of type i can save, so he maximizes utility $u\left(c^y_{t,i}, c^o_{t+1,i}\right)$ subject to the lifetime budget constraint $c^y_{t,i} + c^o_{t,i}R_f^{-1} = W_{t,i}$, where $W_{t,i}$ denotes his net worth at the end of their young period. For a given contract offer $(\delta_{t,i}, R_{t,i})$, wage w_t and capital price ρ_{t+1}, type i will choose whether or not to accept the financial contract offer and engage in production or choose the outside option.

We will drop the time scripts for convenience as we discuss the static aspects of the equilibrium, until reintroducing them in Section 3.4 where equilibrium dynamics is analyzed.

Note that the outside option is to earn wages, which yields $W_i = w$. Because default is only possible at the end of the period, it immediately follows that type B chooses the outside option whenever the early loan payment is unaffordable, $\delta_B < w + \rho f_B$. If the early payment is affordable, type B chooses between the outside option and entering the contract to default, selecting the option that yields the highest net worth:

$$W_B = \max\{w,\ \rho g_B(1 + w + \rho f_B - \delta_B)\}. \tag{1}$$

To understand the above, recall that Assumption 2 ensures that all the proceeds left after making the payment δ_B, that is $1 + w + \rho f_B - \delta_B$, are reinvested into second stage production, which yields capital at the rate g_B. At the end of the period, type B sells his capital at the price ρ and walks away with the proceeds.

Likewise, whenever type G can afford δ_G, it chooses between the outside option and entering the contract to repay. The latter option yields the payoff $\rho(1 + w + \rho f_G - \delta_G)g_G - (2R_G - \delta_G)$ because after selling his capital, type G repays the remaining loan amount $2R_G - \delta_G$. The end of period net worth of type G is given by

$$W_G = \max\{w,\ \rho g_G(1 + w + \rho f_G - \delta_G) - (2R_G - \delta_G)\}. \tag{2}$$

[16] Several of the assumptions involve an endogenous variable, such as ρ_t in this case. These are all made parametric in Section 4.

[17] Intuitively, back-loaded contracts, i.e. those offering $1 - \varepsilon_i$ and $1 + \varepsilon_i$ in the two stages of production, where $\varepsilon_i > 0$, are ruled out immediately as they preclude production. Front-loaded contracts that offer $1 + \varepsilon_i$ in the first stage and $1 - \varepsilon_i$ in the second stage do not benefit type G because second period production is more profitable. However, they make entry more affordable for type B. This tightens the self-selection constraints on G and participation constraint on banks.

[18] If signals are costless, they can be easily mimicked by low productivity entrepreneurs, and therefore contain no informational value. In our context, for example, δ would be costless if banks compensated entrepreneurs for δ with a larger future loan. In this case, there would be scope for mutually beneficial arrangements that would allow low productivity types to borrow money from a short-term lender in order to pay δ. The extra funds that compensate for δ could then be used to pay off the short term lender.

[19] This essentially means that in case of default at the end of the first stage, the bank seizes the entire wealth.

[20] It is straightforward to endogenize the default choice by assuming that banks seize a fraction of output in case of default. The bounds on productivity parameters would then ensure the desired default outcomes: G would choose to default if it were sufficiently productive and B would choose to default if it were sufficiently unproductive. This was done in the previous version of this paper, without adding much insight to the results.

[21] In related literature, the reason why one type of finance is preferred over the other, depending on the state of the economy, is due to the interplay of entrepreneurs' net worth and uncertainty of project payoffs.

Clearly, the optimal choice will depend on the actual productivity parameters and equilibrium prices. In order to illustrate the proposed mechanism, we require that type B is sufficiently productive to choose entry in the case of affordable δ. Note the bound depends on equilibrium prices w and ρ. It is made parametric in Section 4 where the dynamic equilibrium is discussed.

Assumption 3. Type B enters whenever the loan contract is affordable. Assume $\rho g_B \geq w$ whenever $\delta_B \leq \tilde{\delta}$, where $\tilde{\delta}$ denotes the maximum early loan payment affordable by type B,

$$\tilde{\delta}(w, \rho) := w + \rho f_B. \tag{3}$$

In what follows, we proceed to characterize the financial equilibrium. We show that exactly one of the two types of contracts will emerge in equilibrium: either contracts offering the same terms to both types or contracts where type B is excluded from financial markets. In other words, a separating equilibrium will necessarily entail an exclusion of type B from financial markets (Lemma 1). What type of equilibrium emerges will crucially depend on the state of the economy, most importantly, its implications for cash flow constraints faced by young entrepreneurs (Proposition 1).

3.3. Financial market equilibrium

We now characterize the financial market equilibrium for given prices w and ρ. As is known from Rothschild and Stiglitz (1976), an equilibrium may not exist in the context of competitive markets with adverse selection. In short, if separating contracts are preferred by the good type, an equilibrium exists. If, however, pooling contracts are preferred by the good type, competition imposes a zero-profit condition which induces cross-subsidization, and this implies that a bank offering such a contract is exposed to cream-skimming from a competitor targeting the good clients. Hellwig (1987) resolved this dilemma by incorporating a third stage into the game, so that after banks have offered contracts and agents have applied, banks have the possibility to reject their applicants. Using the concept of sequential equilibrium, Hellwig proved that the pooling equilibrium that maximizes the utility of the good type subject to the zero-profit condition can be sustained. Drawing on this result, we follow the standard practice of finding the equilibrium contract pair among the feasible menu as the one that maximizes the utility of type G.

Definition 1. For ρ and w, the financial market equilibrium is given by contracts (δ_G, R_G), (δ_B, R_B) s.t. (1) they induce self-selection and participation by entrepreneurs and participation by banks, (2) banks make zero profits, (3) (δ_G, R_G) is the best loan contract for type G among the contract menu $\{(\delta_G, R_G), (\delta_B, R_B)\}$ satisfying (1) and (2).

The next lemma states that financing of both types cannot occur under distinct contracts. In other words, if distinct contracts are offered to the two types, then the contract offered to type B necessarily excludes him from financial markets. Intuitively, if both types receive financing through different contracts (δ_G, R_G) $\neq$ (δ_B, R_B), self-selection implies that $\delta_B \leq \delta_G$, or else B would choose the contract offered to G. Suppose $\delta_B < \delta_G$. The pooling contract (δ^B, R^G) offered to both types ensures zero profit for the bank because collections from each type are unchanged. Because this contract is strictly preferred by type G, we arrive at a contradiction and conclude that $\delta_B = \delta_G$. The interest rate must also be the same to ensure self-selection by G.[22]

Lemma 1. *If $\delta_B < \tilde{\delta}(w, \rho)$, then $(\delta_B, R_B) = (\delta_G, R_G)$. Otherwise, type B is excluded from financial markets.*

Proof. See the appendix. □

It follows that the only way for type B to be financed is through a pooling contract, and we can restrict attention to contract pairs that either specify identical contracts for both types (a pooling contract), or specify different contracts with non-participation offered to type B. Therefore, it suffices to focus on a single contract (δ, R) offered to type G that either pools or excludes type B. Type B entrepreneurs are excluded by contracts with $\delta \geq \tilde{\delta}(w, \rho)$, and included otherwise. We drop type-dependence at this point to simplify notation.

The zero profit condition pins down the interest rate on any contract characterized by a given $\delta \geq 0$. If $\delta \geq \tilde{\delta}(w, \rho)$, only type G enters, and $R = R_f$. Note that type G will strictly prefer $\left(\tilde{\delta}(w, \rho), R_f\right)$ to any other separating contract because his payoff decreases in δ. This is seen clearly from (2): cash flows not spent on δ are reinvested and yield a positive return in stage two production (Assumption 2). Hence, we can eliminate contracts with $\delta > \tilde{\delta}(w, \rho)$ as potential equilibrium outcomes.

For $\delta < \tilde{\delta}(w, \rho)$, type B prefers to enter production. Because types are unobservable, loanable funds (2μ in total) are disbursed according to the distribution of types in the population, with each entrant obtaining 2 units of funds in total. Hence, a proportion μ^2 of entrepreneurs of type G and a proportion $(1-\mu)\mu$ of type B are financed. Loan payments collected consist of full repayment from type G entrepreneurs, $2\mu R$, and the early payment $(1-\mu)\mu\delta$ collected from type B. The opportunity cost of loaning out funds is investment into risk free technology, $2\mu R_f$. Hence, the interest rate on this pooling contract is the solution to the zero profit condition:

$$\mu^2 2R + (1-\mu)\mu\delta = 2\mu R_f. \tag{4}$$

Therefore, the interest rate consistent with the zero profit condition is summarized by

$$R(w, \rho, \delta) = \begin{cases} \dfrac{R_f}{\mu} - \dfrac{(1-\mu)}{2\mu}\delta & \text{if } \delta < \tilde{\delta}(w, \rho) \\ R_f, & \text{if } \delta = \tilde{\delta}(w, \rho) \end{cases}, \tag{5}$$

where we solved (4) for R. Note that $R(w, \rho, \delta)$ decreases in δ in the interval $[0, \tilde{\delta}(w, \rho))$ where both types are financed. This is because δ is collected early, prior to default, and therefore an increase in δ raises the banks' total collection from type B and thus requires a lower cross-subsidy from type G.

Therefore, among the contracts satisfying (1) and (2) from the equilibrium definition, the relevant set is of the form

$$\Omega(w, \rho) \equiv \left\{(\delta, R(w, \rho, \delta)) | \delta \leq \tilde{\delta}(w, \rho), G \text{ enters}\right\}.$$

Assumption 4(a), stated below, ensures that an increase in δ indeed lowers the payoff to type G, i.e. that our model captures the idea that, in loan markets where information is hard to gather, credible signaling is costly, and the cost is borne by the productive type. It requires that investment in second stage production must be sufficiently profitable so that the net worth reduction implied by the direct effect of increasing δ (which effectively constrains reinvestment in second stage production) is not offset by the gain in net worth due to the lower interest rate (the indirect effect). Precisely, this assumption is obtained by setting $dW_G/d\delta < 0$ and simplifying.[23] This assumption effectively establishes a lower bound on type G's productivity. Assumption 4(b) ensures G's entry and therefore guarantees that the set $\Omega(w, \rho)$ is non-empty.

[22] Technically speaking, it is possible to have $R^B \geq R^G$ in equilibrium. Because type B does not care about the interest rate, we restrict attention to equality without any loss of generality.

[23] Using the payoff definition (2), we obtain $\frac{dW_G}{d\delta} = -\rho g_G - 2\frac{\partial R}{\partial \delta} + 1$ for the case of pooling, where $\frac{\partial R}{\partial \delta} = -\frac{1-\mu}{2\mu}$ as seen from (5). Setting $\frac{dW_G}{d\delta} < 1$ and simplifying gives $\mu\rho g_G > 1$.

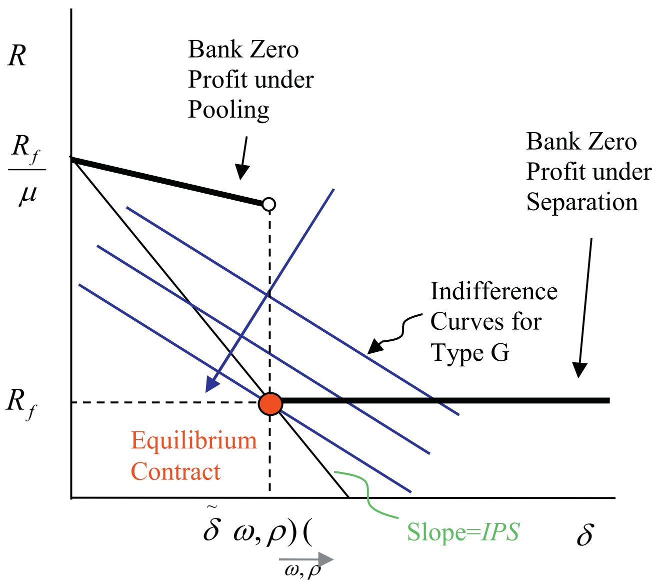

Fig. 3. Equilibrium contract determination (for given w and ρ). *Notes*: this figure illustrates the determination of the financial contract as a contract that maximizes utility of type Gentrepreneur subject to the banks' zero profit condition. The indifference curves for type Gentrepreneur indicate that utility decreases in the interest rate R and the early loan payment δ . In this particular case, the utility is maximized by offering a separating contract, which requires a high early loan payment $\tilde{\delta}$(in order to exclude type B) but offers a low interest rate. In this case, the slope of the iso-profit line (IPS) is steeper than the slope of the indifference curves.

Assumption 4. Suppose the following conditions hold for prices $w,\ \rho$: a). Signaling is costly: $\mu\rho g_G > 1$, b). Type G participates: $\mu\rho g_G + \rho f_B \geq 2R_f$.

The following lemma shows that the set $\Omega(w,\ \rho)$ is indeed nonempty. Whether the contract offers pool the two types or separate them, type G enters the financial contract and repays.

Lemma 2. *The following results hold:*

1. For any $\delta < \tilde{\delta}(w, \rho)$, the pooling contract ($\delta$, $R(w,\ \rho,\ \delta)$) induces G's entry. Moreover, $R(w,\ \rho,\ \delta) > R_f$.

2. The separating contract $(\tilde{\delta}(w, \rho), R_f)$ induces G's entry.

Therefore, we have characterized the set of loan contracts that induce self-selection and participation by entrepreneurs and banks and ensure that banks make zero profits. According to the definition of the financial market equilibrium, it remains to identify which contract among these maximizes the payoff for type G.

Under any contract, for given prices ($w,\ \rho$), the net worth obtained by type G is given by $\rho(g_G + (w + \rho f_G - \delta)g_G) - (2R - \delta)$. Therefore, the equilibrium contract offered by banks is a solution to

$$\max_{(\delta, R)} \quad \rho(g_G + (w + \rho f_G - \delta)g_G) - (2R - \delta)$$

$$\text{s.t.} \quad R = R(w, \rho, \delta),$$

where $R(w,\ \rho,\ \delta)$ is given by Eq. (5).

This problem, which yields the equilibrium financial contract, is summarized by the feasible set (zero-profit condition) and type G indifference curves, depicted in Fig. 3. Indifference curves describing his trade-off between R and δ, given current prices, have the slope

$$ICS := -\frac{\rho g_G - 1}{2} < 0, \tag{6}$$

which is negative by Assumption 2. Moreover, the indifference curves become steeper as ρ rises. Intuitively, as capital sells for a higher price, the cost of foregone investment associated with δ rises, and type G requires a greater reduction in R to compensate for a given rise in δ. The solution depends on the relative size of *ICS* and the slope of the line connecting the following two points on the lender's zero-profit curve: the pooling contract $(\delta, R) = (0, R(w, \rho, 0))$ and the separating contract $(\tilde{\delta}(w, \rho), R_f)$. We call this slope *IPS*, the first two letters referring to the iso-profit,

$$IPS := -\frac{R(w, \rho, 0) - R_f}{\tilde{\delta}(w, \rho)} < 0. \tag{7}$$

Fig. 3 depicts the solution to this linear banking problem: the pooling contract (0, $R(w,\ \rho,\ 0)$) is offered if $ICS < IPS$, and the separating contract $(\tilde{\delta}(w, \rho), R_f)$ is offered otherwise. In the figure, the indifference curves are drawn relatively flat, which leads to a separating equilibrium. The figure also clarifies that in times of high cash flows (high w and ρ), $\tilde{\delta}(w, \rho)$ would increase making screening more expensive for type G . As a result, IPS would flatten and may lead to a pooling equilibrium where screening is completely abandoned.

Under separation, total capital production is carried out by measure μ of type G entrepreneurs:

$$\begin{aligned} k_s(\rho) &= \mu\big(f_G + g_G + (\rho f_G + w - \tilde{\delta}(w, \rho))g_G\big) \\ &= \mu(f_G + g_G + (\rho(f_G - f_B))g_G). \end{aligned} \tag{8}$$

Under the pooling contract, measure μ^2 of type G entrepreneurs and measure $\mu(1-\mu)$ of type B entrepreneurs obtain financing and enter production. Since the production of type i under the contract that involves no early payment is given by $f_i + g_i + (\rho f_i + w)g_i$, the total capital production is given by

$$\begin{aligned} k_p(w, \rho) = \mu^2&(f_G + g_G + (\rho f_G + w)g_G) \\ &+ \mu(1-\mu)(f_B + g_B + (\rho f_B + w)g_B). \end{aligned} \tag{9}$$

The following proposition summarizes these results.

Proposition 1. *Financial Equilibrium, for given prices. If $\frac{1-\rho g_G}{2} \geq -\frac{R(w,\rho,0)-R_f}{\tilde{\delta}(w,\rho)}$, then a separating contract $(\tilde{\delta}(w, \rho), R_f)$ is offered, and total capital production is given by (8). Otherwise, a pooling contract* (0, $R(w,\ \rho,\ 0)$) *is offered, and total capital production is given by (9).*

Proof. Follows from the above discussion. □

3.4. Dynamic general equilibrium

We characterized the financial market equilibrium and derived capital production for given prices. As usual, prices are determined endogenously, but in this case the consumption good firms producing in period $t+1$ must also form beliefs in period t about wages w_{t+1}, which are not announced in period t but influence the decision about the demand for capital k_{t+1}. These beliefs must be consistent with the actual wages, which in turn must clear the labor market in period $t+1$, given the total capital k_{t+1} purchased by the consumption good sector.

An expectation about future wages can then be seen as an expectation about the future level of capital k_{t+1}, which must be consistent with the actual decisions of entrepreneurs and firms in period t.

Definition 2. A dynamic equilibrium for given k_0 is given by sequences of prices $\{w_t^*, \rho_t^*\}_{t=0}^{\infty}$, capital levels $\{k_{t+1}^*\}_{t=0}^{\infty}$, beliefs $\{k_t'^*\}_{t=0}^{\infty}$ and contracts $\{R_t^*, \delta_t^*\}_{t=0}^{\infty}$ such that the following holds:

- Expectations are rational, i.e. consistent with equilibrium outcomes: $k_t'^* = k_{t+1}^*$.
- Prices $w_t^* = A_t(1-\beta)k_t^{*\beta}$ and $\rho_{t+1}^* = A_{t+1}\beta\big(k_t'^*\big)^{\beta-1}/R_f$ satisfy Assumptions 3 and 4.

- Contracts and capital production, for given prices, are determined as in Proposition 1:

$$(\delta_t^*, R_t^*) = \begin{cases} (\tilde{\delta}(w_t^*, \rho_{t+1}^*), R_f) & \text{if } \frac{1-\rho_{t+1}^* g_G}{2} \geq -\frac{R(w_t^*, \rho_{t+1}^*, 0) - R_f}{\tilde{\delta}(w_t^*, \rho_{t+1}^*)} \\ (0, R(w_t^*, \rho_{t+1}^*, 0)) & \text{otherwise} \end{cases},$$

where $\tilde{\delta}(w, \rho)$ is defined in (3) and $R(w_t^*, \rho_t^*, 0)$ is defined in (5);

$$k_{t+1}^* = \begin{cases} k_s(\rho_{t+1}^*) & \text{if } \frac{1-\rho_{t+1}^* g_G}{2} \geq -\frac{R(w_t^*, \rho_{t+1}^*, 0) - R_f}{\tilde{\delta}(w_t^*, \rho_{t+1}^*)} \\ k_p(w_t^*, \rho_{t+1}^*) & \text{otherwise} \end{cases},$$

where $k_s(\rho)$ and $k_p(w, \rho)$ are given by (8) and (9).

4. Equilibrium characterization with constant capital prices

In order to effectively illustrate the insights offered by our model, we assume the presence of a productivity externality in consumption good production due to the size of the economy: $A_t = K_t^{1-\beta}$. This assumption ensures that the price of capital is constant, while the wage w_t varies linearly with the capital level k_t:

$$\rho_t = \beta / R_f \text{ and } w_t = (1-\beta) k_t. \tag{10}$$

Because the price of capital is constant and cash flows change only due to the labor income variation, k_t is the only state variable and expectations about future capital do not play any role. The solution and intuition are easy to obtain in this case, without compromising much insight. Higher wages raise the level of the early payment needed for effective screening, implying that, for high states of the economy, screening is abandoned, and pooling contracts emerge in equilibrium. The composition of entrepreneurs deteriorates as a result, bringing about sufficiently low levels of capital and wages that deem screening inexpensive, and therefore leading to separating contracts.

The consumption good output production simplifies to $F(K_t, L_t, t) = k_t$. Because capital depreciates upon use by the consumption good firm, we think of it as an intermediate good and do not count it in the model measure of total output.[24] It is then clear that capital, final good (i.e. consumption good) output, its productivity (A_t) and wages will all move together, so it suffices to focus our discussion of model dynamics on k_t alone.

To simplify, let $R_f = 1$. Drawing on the earlier derivation of equilibrium contracts for given prices, we first obtain equilibrium contracts as a function of the state variable k_t. The minimum early payment that accomplishes separation becomes

$$\tilde{\delta}(k_t) = (1-\beta) k_t + \beta f_B$$

and the maximization problem that determines the equilibrium contract, for a given k_t, simplifies to

$$\max_{(\delta, R)} \ \rho(g_G + ((1-\beta)k_t + \beta f_G - \delta) g_G) - (2R - \delta)$$

$$\text{s.t. } R = R(k_t, \delta) = \begin{cases} \frac{R_f}{\mu} - \frac{(1-\mu)}{2\mu}[\delta + (1-\alpha)\beta(g_B + ((1-\beta)k_t + \beta f_B - \delta) g_B)], & \text{if } \delta < \tilde{\delta}(k_t) \\ R_f, & \text{if } \delta = \tilde{\delta}(k_t) \end{cases}.$$

Expressions (6) and (7) now become $ICS = -\frac{\beta g_G - 1}{2}$ and $IPS = -\frac{R(k_t, 0) - R_f}{\tilde{\delta}(k_t)}$. Proposition 1, which characterized contracts offered in equilibrium, can be now expressed in terms of k_t. If $\frac{1-\beta g_G}{2} \geq -\frac{R(k_t, 0) - R_f}{\tilde{\delta}(k_t)}$, then a separating contract $(\tilde{\delta}(k_t), R_f)$ is offered. Otherwise, a pooling contract (0, $R(k_t, 0)$) is offered.

Fig. 4 illustrates the equilibrium contract determination for two different levels of capital, low and high, and clarifies the main intuition for the link between the state of the economy and stringency of lending standards. Note that *ICS* is independent of k_t, while *IPS* increases in k_t, thus leading to the selection of separating contracts for low levels of k_t and pooling contracts for high levels of k_t. Intuitively, cash flow increases with the state of the economy, thereby improving the ability of low productivity types to make loan payments and raising the payment $\tilde{\delta}(k_t)$ that keeps type *B* out. This raises the cost of effective screening, captured as an endogenous restriction on type *G*'s investment imposed by $\tilde{\delta}(k_t)$. Therefore, when economic fundamentals are strong and cash flow is high, separation is particularly unattractive to *G* as it requires a very expensive signal.

There is a threshold $\bar{k}$ dividing the state space into regions of pooling and separating contracts.

Lemma 3. *Define* $\bar{k} = \frac{(1-\mu)}{\mu(1-\beta)} \frac{2R_f}{\beta g_G - 1} - \frac{\beta}{1-\beta} f_B$. *For* $k_t > \bar{k}$, *a pooling contract is selected, while for* $k_t \leq \bar{k}$, *a separating contract is selected.*

Proof. The result follows immediately from $\bar{k}$ being the unique solution to $\frac{\beta g_G - 1}{2} = \frac{R(k_t, 0) - R_f}{\tilde{\delta}(k_t)}$. □

We explained how changes in cash flow help interpret the empirical link between the state of the economy and the stringency of lending standards. We now discuss the implications for productivity dynamics. Screening out the bad projects becomes more costly in expansions, which leads to the selection of pooling contracts. This outcome is suboptimal from the point of view of the planner facing the same informational constraints. Low productivity entrepreneurs enter production alongside the productive type, the composition effect setting off a recession. This situation, in turn, eventually generates conditions – a low enough level of capital and cash flow – conducive to the emergence of separating contracts that exclude low productivity entrepreneurs from production. Optimality is restored, higher productivity in the capital sector leading to an economic recovery.

The time t capital production, characterized in Proposition 1, can be written as a function of k_t:

$$k_s = \mu(f_G + g_G + \beta(f_G - f_B) g_G),$$

$$k_p(k_t) = \mu^2(f_G + g_G + ((1-\beta)k_t + \beta f_G) g_G) + (1-\mu)\mu(f_G + g_B + ((1-\beta)k_t + \beta f_B) g_B).$$

Note that k_s is independent of capital because an additional unit of capital translates into $1-\beta$ additional units of labor income, which are paid to the bank before the second stage takes place in order to keep separation viable, and therefore they are not reinvested. However, $\frac{dk_p(k_t)}{dk_t} > 0$ because a pooling contract entails no early payment and therefore every additional unit of labor income is reinvested, augmenting current capital production. Invoking Lemma 3 we can then derive the transition function for capital as

$$k_{t+1} = \begin{cases} k_s & \text{if } k_t \leq \bar{k} \\ k_p(k_t) & \text{otherwise} \end{cases}. \tag{11}$$

How does k_s relate to $k_p(k_t)$ for any given level of k_t? On the one hand, under separation, all of the productive entrepreneurs engage in production. Type *B* entrepreneurs do not apply for financ-

[24] One could include it without changing any important insights, but the exposition would get more tedious.

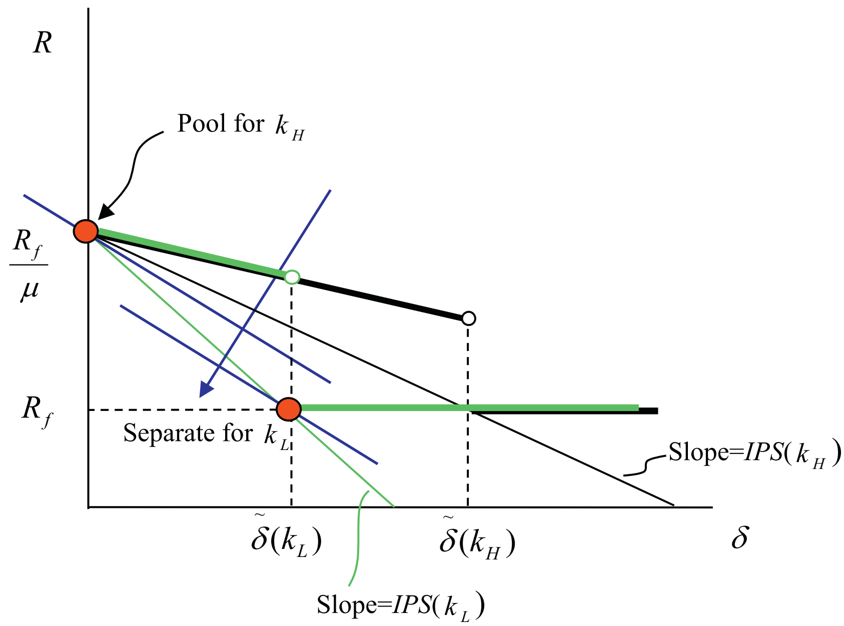

Fig. 4. Equilibrium contract determination for two values of the aggregate state, $k_H > k_L$. *Notes*: this figure illustrates the determination of the financial contract under two different states of the economy for the model with production externality, analyzed in Section 4. When capital is high (k_H), cash flow is also high. This means that the early payment that excludes type B , i.e. $\tilde{\delta}(k_H)$, is also high. It is clearly seen that, in this case, utility of type G is maximized via a pooling contract. Type G prefers a contract with no early repayment which allows for high levels of investment, even though it requires a higher interest rate. The opposite happens when the aggregate state is low (k_L). Because cash flow is also low, exclusion of type B is viewed as relatively cheap by typeGas it entails a low required early payment $\tilde{\delta}(k_L)$. A separating contract arises in this case.

ing and do not crowd out type G.[25] Therefore, selection of pooling over separation for any state k entails this negative composition effect at the extensive margin. On the other hand, under separation, screening necessarily restricts investment in capital projects and therefore each type G entrepreneur produces less. Therefore, selection of pooling over separation entails a positive investment effect at the intensive margin. To illustrate the productivity reversion mechanism, we focus on the set of parameters for which the negative extensive margin dominates the positive intensive margin at $\bar{k}$.

Assumption 5. Jump in the Transition Function. $k_p(\bar{k}) < k_s$.

We also make an assumption to ensure there is no perpetual growth in this economy: under pooling, an extra unit of capital, which translates into an extra $1-\beta$ units of input into the second stage of production, results in less than one unit of additional capital.

Assumption 6. No Perpetual Growth. $\frac{dk_p(k_t)}{dk_t} = (1-\beta)\mu(\mu g_G + (1-\mu)g_B) < 1$.

The proposition below states restrictions on the set of parameters and initial condition k_0 that ensure existence of a dynamic general equilibrium that exhibits cyclical behavior. It also ensures that Assumptions 3 and 4 hold for all possible values of k_t. We focus on the existence of cyclical equilibria, because they demonstrate the productivity reversal mechanism in the most stark manner.

Proposition 2. *Cyclical Dynamics. Suppose that* $k_p(\bar{k}) < \bar{k} < k_s$. *Define* $k_{\min} := k_p(\bar{k})$ *and* $k_{\max} := k_s$. *Suppose that Assumptions 1–6 hold for* $\rho = \beta$ *and* $w = (1-\beta)k_{\max}$. *Then for any* $k_0 \in [k_{\min}, k_{\max}]$, *there exists a dynamic equilibrium with the capital stock (and output) exhibiting cycles, not necessarily trivial (see Fig. 5), the socially optimal allocation of funds attained only at the very bottom of the cycle.*

Proof. See the Appendix. □

An economic expansion leads to the endogenous relaxation of lending standards, which allows for financing of the less productive entrepreneurs and consequently negatively impacts future productivity through changing the composition of entrepreneurs.

From the point of view of a social planner with the same informational constraints, pooling regime is suboptimal along the entire equilibrium path because $k_s > k_p(k)$ for all k. Note that the planner cannot observe individual types and therefore cannot improve the outcome beyond enforcing the financial regime that maximizes capital production for a given k. The planner would always enforce $\delta = \tilde{\delta}(k)$.[26] It follows that the socially optimal outcome is achieved only at the very bottom of the cycle. At all other points of the cycle, funds are misallocated. The insight offered by our model is that competition for the productive types in times of high cash flow, and therefore particularly high costs of sending an effective signal (screening costs), leads to contract offers that entail no screening – lemons enter as a side effect.

For cyclical equilibria, the length of the cycle can be easily calculated from the primitives.[27]

Corollary 1. *Consider an economy satisfying* $k_p(\bar{k}) < \bar{k} < k_s$. *If* n *is the smallest number such that* $k_p^{(n-1)}(k_s) > \bar{k}$ *but* $k_p^n(k_s) < \bar{k}$, *then an economy starting at* $k_0 = k_s$ *exhibits cycles of length* $n+1$. *In each*

[25] Crowding out of type G caused by financing of type B entrepreneurs occurs because there is a limited supply of funds. This, however, is not essential for our results. What is essential is that financing of type B entrepreneurs crowds out investment into any technology with higher returns.

[26] An attempt to concentrate production among a smaller measure of type G entrepreneurs would result in the entry of the less productive types.

[27] To complete the full description of possible equilibrium dynamics, observe that it is also possible that parameters are such that $\bar{k} < k_p(\bar{k}) < k_s$ and the capital stock converges to $k_{ss} = k_p(\bar{k})$. Finally, another possible equilibrium behavior is for the capital stock to converge to k_s. Such behavior would arise for parametric restrictions that guarantee existence of equilibrium and ensure that $k_p(\bar{k}) < k_s < \bar{k}$.

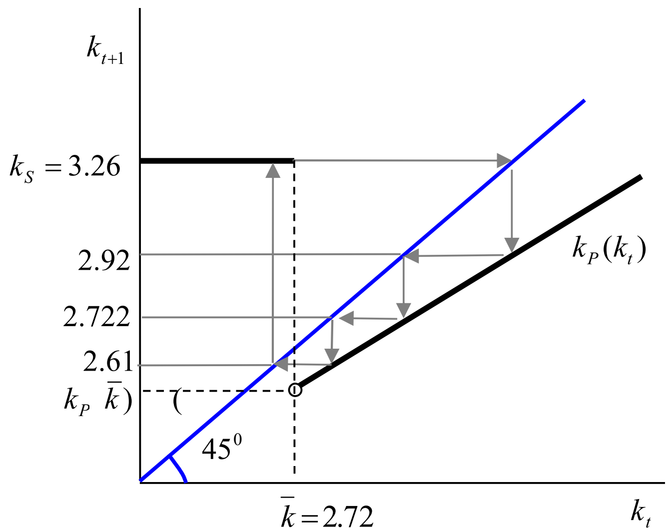

Fig. 5. Transition function, $k_{t+1}(k_t)$. *Notes*: this figure illustrates the transition function, k_{t+1}as a function of k_t for the model with production externality, analyzed in Section 4. It illustrates the possibility of endogenous cycles in the case of $k_P(\bar{k}) < \bar{k} < k_S$. Example 1 is the corresponding numerical example that produces the transition function and dynamics illustrated in this figure. At the trough, when capital is at its lowest, so are the final good output, productivity and wages. Low cash flows imply that screening is sufficiently cheap so that lending standards get tight, i.e. type B is excluded from production. The strong entrepreneur composition effect (positive extensive margin) ensures that more capital becomes available to the aggregate firm after these entrepreneurs finish their projects. In fact, capital next period is at its highest possible level, k_S. As a result, productivity, output and wage payments soar. In turn, high cash flows deem signaling expensive and bring about a relaxation of screening standards, allowing for the entry of type B entrepreneurs. The worsening of entrepreneur composition (negative extensive margin) sets off a recession. Investment in capital production projects falls for the entire duration of the downturn because of declining wages.

cycle, the capital level declines for the first $n-1$ periods and goes up to k_s in the last period.

Proof. See the appendix. □

We provide a numerical example to demonstrate the possibility of non-trivial cycles depicted in Fig. 5. We set β to 0.4 to match the labor income share in the U.S. The risk-free rate is already set to 1. We choose very low productivity in the initial production stage to capture the idea that second stage production is very productive which is the essence of costly signaling. We then choose entrepreneur productivity parameters in second stage so that the relative average productivity of type G and type B approximate the relative size of small businesses surviving the 5 year mark and those that do not. That is approximately a factor of 4.9.[28] Finally, we set μ low enough to make sure that when type B entrepreneurs enter, they send the economy in a recession.

Example 1. (Fig. 6) *Cyclical Dynamics. Assume* $\beta = 0.4,\ R_f = 1,\ f_G = 0.2,\ f_B = 0,\ g_G = 23,\ g_B = 5.2,\ \mu = 0.13$. *With these parameters,* $\bar{k} = 2.72,\ k_s = 3.26,\ k_p(k_s) = 2.9$. *Because these parameters imply* $k_p(k_s) > \bar{k}$, *we obtain 4-period cycles (3.26, 2.92, 2.722, 2.61, 3.26,...).*

This example effectively illustrates how endogenous changes in financial arrangements reverse productivity trends through the extensive margin (i.e. composition of financed entrepreneurs). At the trough, when capital is at its lowest, so are the final good output, wages and productivity $\left(A_t = k_t^{1-\beta}\right)$. Low wages imply that signaling is sufficiently cheap for screening standards to return to the

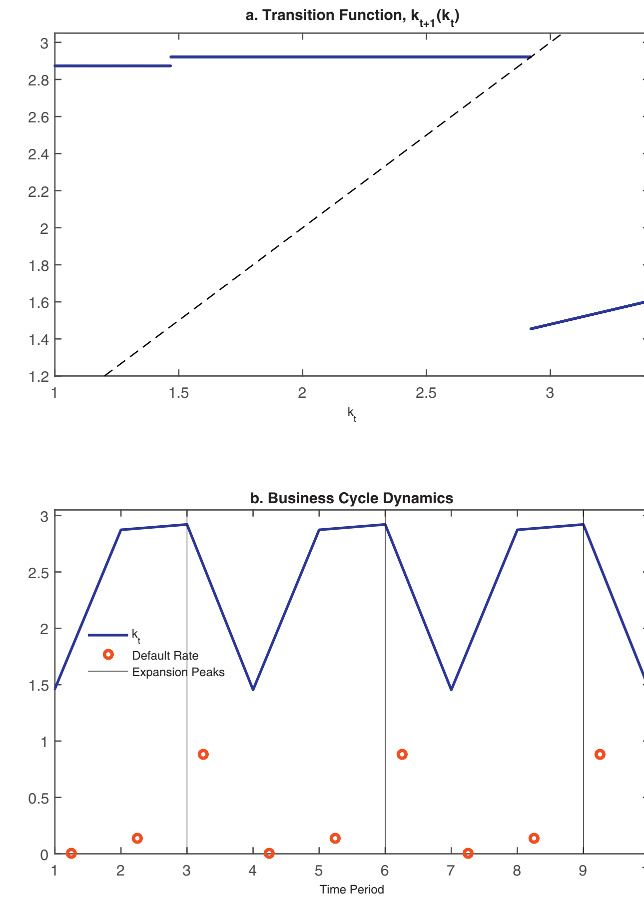

Fig. 6. Transition function $k_{t+1}(k_t)$ and cyclical dynamics with three types of entrepreneurs. Notes: this figure presents the transition function and cyclical dynamics for the model extended to include three types of entrepreneurs. The corresponding numerical example that generates the graphs is Example 2 in Section 4. When k is low only type G enters production which raises capital to a medium level. Screening standards weaken which allows for greater investment but type M enters production. Although type M generates default, capital increases further, to a high level, due to greater investment rates. Screening standards relax further, allowing for type B entrepreneurs to enter production. Because the measure of low productivity entrepreneurs is high, the negative composition effect dominates the positive investment effect, resulting in low capital production. The cyclical behavior is then repeated.

optimal level, which ensures that bank funds are allocated to the best entrepreneurs (positive extensive margin). This captures the cleansing effect of recessions. Despite the low investment rates in their projects (negative intensive margin),[29] more capital becomes available to the aggregate firm after these entrepreneurs finish their projects. As a result, its productivity, output and wage payments soar as the economy peaks. In turn, high wages deem signaling expensive and bring about a pooling regime that allows for greater reinvestment rates (positive intensive margin) but worsens the composition of entrepreneurs (negative extensive margin). The extensive margin dominates again, setting off a recession. Investment in capital production projects falls for the entire duration of the downturn as wages decline while screening standards remain lax.[30]

The model described thus far produces expansions that last only one period, which makes it difficult to assess the insight

[28] We obtain firm size information from https://www.sba.gov/advocacy/firm-size-data.

[29] Investment rates are low because cash flows are used in signaling.

[30] In the case of pooling, no resources are spent on signaling and therefore all cash flows are invested in capital projects, in addition to the 2 units obtained from the bank. Total investment then equals $\mu^2(2 + w_t + \beta f_G) + (1-\mu)\mu(2 + w_t + \beta f_B)$.

our mechanism offers for variables that change *during* expansions. We extended our model in a straightforward way to include three types of entrepreneurs: G, M and B. Type M is more productive than type B in both stages, but he still prefers default. This generalization more clearly delineates periods of expansions and recessions, and it is our preferred version of the model for illustrating the comovement of macroeconomic variables over the business cycle. This setup also allows for longer lasting expansions and therefore helps us explain the dynamics of default rates observed in the data, that is, that default rates begin to rise while the economy is still expanding, peaking shortly after the recessions set on.

The parametric example given below and illustrated in Fig. 6 delivers a three-period cycle characterized by the following dynamics. For the lowest k, only type G is financed and hence default rates are zero. The economy grows due to the strong composition effect, and despite the low investment rates of type G entrepreneurs. Next period, lending standards are relaxed allowing the less productive type M to be financed along with type G. The economy, however, continues to grow as the increase in reinvestment along the intensive margin dominates the negative composition effect along the extensive margin. Default rates are now positive as the type M entrepreneurs choose to default. Next period, lending standards are relaxed further to a point where the worst entrepreneurs are financed along with G and M. Despite the high levels of investment, the extensive margin dominates at this point of the cycle, implying a large drop in productivity. Default rates peak shortly after recessions set in, as both type B and M entrepreneurs choose to default. Importantly, this example shows that it is possible for the amplification and reversion mechanisms, reviewed in Section 2, to coexist, with their relative strength varying over the business cycles.

This straightforward extension of the model allows us to identify periods of expansions and recessions more clearly. The model dynamics is qualitatively consistent with the dynamics of net worth, cash flows, investment, average loan size, default rates and entrants' quality over the business cycle. It can even account for the fact that expansions usually last longer than recessions in the data.

Example 2. (Fig. 6[31]) *Cyclical Dynamics with Three Types. Assume* $\beta = 0.4$, $R_f = 1$, $f_G = 0.1$, $f_M = 0.05$, $f_B = 0$, $g_G = g_M = 24.4$, $g_B = 1.5$, $\mu_G = 0.12$, $\mu_B = 0.02$. *With these parameters, we obtain a 3-period cycle. When k is low only type Genters production which raises capital to a medium level. Screening standards weaken which allows for greater investment but type M enters production. Although type M generates default, capital increases further, to a high level, due to greater investment rates. Screening standards relax further, allowing for type B entrepreneurs to enter production. Because the measure of low productivity entrepreneurs is high, the negative composition effect dominates the positive investment effect, resulting in low capital production. The cyclical behavior is then repeated.*

The rest of the paper generalizes the model in several ways, without altering the important insights of the mechanism at work.

5. Endogenous loanable funds

In this section, we extend the analysis of the model characterized in the previous section by endogenizing the supply of loanable funds, previously fixed at 2μ. Precisely, we require that loans to the young entrepreneurs are financed with the old entrepreneurs' savings. To keep the analysis tractable, we keep the risk-free interest rate fixed as in the case of a small open economy.

Endogenizing the supply of loanable funds not only confirms the possibility of cyclical economic behavior, but it also gives rise to the possibility of sudden drops and slow recoveries. It also generates predictions that are qualitatively consistent with the relevant empirical evidence: procyclicality of net worth, cash flows, investment and loanable funds, as well as default rates lagging after the business cycle.

The intuition for slow expansions and sudden recessions is as follows. As discussed in Section 4, for low enough levels of capital, separating contracts emerge and only type G entrepreneurs enter production. However, if the supply of funds is also low, only a small fraction of type G entrepreneurs is financed. Aggregate productivity, capital production and net worth of active entrepreneurs are high. As a result, savings of active entrepreneurs are high, contributing positively to the supply of loanable funds next period. Of course, there are entrepreneurs that are either screened out or rationed out of financial markets, whose income and savings are low. If the positive effect on aggregate savings is sufficiently strong, the supply of loanable funds rises. A greater measure of type G entrepreneurs get financed and the recovery continues until capital reaches a high enough level that induces the selection of pooling contracts. As pooling emerges and type B entrepreneurs enter production, the high levels of funds and investment may not offset the decline in productivity due to a worsening in the mix of entrepreneurs engaged in production. Hence, the model gives rise to cyclical dynamics that exhibit long expansions and sudden recessions.

In what follows, we derive the dynamical system describing the evolution of the two state variables (capital and savings). We then analyze their behavior using the phase diagram and illustrate the possibility of cyclical dynamic equilibria exhibiting sudden drops and slow (possibly non-monotone) recoveries.

5.1. Transition functions

The funds used to finance projects implemented by the current young are given by the savings S_t of the current old.[32] If these funds are not sufficient to finance all applicants, crowding out occurs, with the unfinanced entrepreneurs staying out of capital production and obtaining only $w(k_t)$ as labor income. If funds are in excess of applicants' demand, every applicant gets financed. Excess funds are saved in international markets at the risk-free rate R_f.

It follows from the discussion in Section 4 that whether pooling or separating contracts emerge depends on k_t (according to Lemma 3) and not on S. The amount of savings S, however, is important as it affects the measure of entrepreneurs financed and influences capital and loanable funds in the next period. In case of a pooling contract, the available funds are used to finance both type G and type B entrepreneurs, with the total funds demanded in the amount of 2. In the case of separation, only type G is financed, and the total demand for funds equals 2μ. Capital stock in period $t+1$, given the current state variables k_t, S_t, is then

$$k_{t+1}(S_t, k_t) = \begin{cases} \min\{\frac{S_t}{2}, 1\}\left[\mu k_p^G(k_t) + (1-\mu)k_p^B(k_t)\right] & \text{if } k_t > \bar{k} \\ \min\{\frac{S_t}{2}, \mu\}k_s^G & \text{if } k_t \le \bar{k} \end{cases}, \tag{12}$$

where $k_s^G = f_G + g_G + \beta(f_G - f_B)g_G$ and $k_p^i(k_t) = f_i + g_i + ((1-\beta)k_t + \beta f_i)g_i$ denote individual capital production of type G entrepreneur under separation and of each type under pooling, respectively.

Homotheticity of $u(c^y, c^o)$ and constant R_f imply that entrepreneurs always save a constant fraction $(:= \xi)$ of their net

[31] In the figure, we mark midpoints of time periods. Recall that default rates occur in the second stage.

[32] The current old collect the gross return to their savings at the end of the period, consuming the entire amount.

worth at the end of their young period. Similarly, the unfinanced entrepreneurs save the same fraction ξ of their wage income $(1-\beta)k$. We obtain the supply of funds tomorrow as a function of the current state variables:

$$S_{t+1}(S_t, k_t) = \begin{cases} \xi\left[\frac{S_t}{2}\left(\mu W_p^G(k_t) + (1-\mu)W_p^B(k_t)\right) + \left(1-\frac{S_t}{2}\right)(1-\beta)k_t\right] & \text{if } k_t > \bar{k} \text{ and } S_t \le 2, \\ \xi\left[\mu W_p^G(k_t) + (1-\mu)W_p^B(k_t)\right] & \text{if } k_t > \bar{k} \text{ and } S_t > 2, \\ \xi\left[\frac{S_t}{2}W_s^G + \left(1-\frac{S_t}{2}\right)(1-\beta)k_t\right] & \text{if } k_t \le \bar{k} \text{ and } S_t \le 2\mu, \\ \xi\left[\mu W_s^G + (1-\mu)(1-\beta)k_t\right] & \text{if } k_t \le \bar{k} \text{ and } S_t > 2\mu, \end{cases} \tag{13}$$

where the net worth of type G young entrepreneur under separation, type G young entrepreneur under pooling, and type B young entrepreneur under pooling are given by

$$W_s^G = \beta[g_G + \beta(f_G - f_B)g_G] - 2R_f,$$

$$W_p^G(k_t) = \beta[g_G + ((1-\beta)k_t + \beta f_G)g_G] - \frac{2R_f}{\mu},$$

$$W_p^B(k_t) = \beta[g_B + ((1-\beta)k_t + \beta f_B)g_B].$$

The definition of the dynamic equilibrium (Definition 2) must be modified for the extended environment. In particular, the equilibrium path of $\{S_t^*\}_{t=0}^{\infty}$ must be specified. While the prices and contracts are determined as previously, k_t^* and S_t^* now evolve according to (12) and (13). Note that prices must still satisfy Assumption 3.

5.2. The phase diagram

As usual for the analysis of such a dynamical system, we divide the state space (S_t, k_t) into regions where it is possible to sign the changes $k_{t+1} - k_t$ and $S_{t+1} - S_t$. Because only k_t matters for whether separating or pooling contracts emerge, the horizontal line given by $k_t = \bar{k}$ splits the state space (S_t, k_t) into the region of separating contracts (below the line) and the region of pooling contracts (above).

First, consider the region of separation ($k_t \le \bar{k}$). From (12) and (13), we have $k_{t+1} = k_t$ iff

$$k_t = \min\left\{\frac{S_t}{2}, \mu\right\}k_s^G \tag{14}$$

and $S_{t+1} = S_t$ iff

$$k_t = \begin{cases} \dfrac{(2/\xi - W_G^s)S_t}{(1-\beta)(2-S_t)} & \text{if } S_t \le 2\mu \\ \dfrac{S_t/\xi - \mu W_G^s}{(1-\beta)(1-\mu)} & \text{if } S_t > 2\mu \end{cases}. \tag{15}$$

Note that (14) is an upward sloping linear curve for $S_t \le 2\mu$, at which point it connects to a horizontal line. The change in capital stock $k_{t+1} - k_t$ is negative for points above and to the left of (14), and positive otherwise. Moreover, the change in savings $S_{t+1} - S_t$ is positive above (15) and negative below it.

In the pooling region ($k_t > \bar{k}$), we obtain $k_{t+1} = k_t$ iff

$$k_t = \frac{A' \min\{S_t, 2\}}{2 - B' \min\{S_t, 2\}}, \tag{16}$$

where $A' = \mu(f_G + g_G + \beta f_G g_G) + (1-\mu)(f_B + g_B + \beta f_B g_B)$ and $B' = \mu(1-\beta)g_G + (1-\mu)(1-\beta)g_B$, and we also obtain $S_{t+1} = S_t$ iff

$$k_t = \begin{cases} \dfrac{(2/\xi - A)S_t}{BS_t + (2-S_t)(1-\beta)} & \text{if } S_t \le 2 \\ \dfrac{S_t/\xi - A}{B} & \text{if } S_t > 2 \end{cases}, \tag{17}$$

where $A = \beta[\mu(g_G + \beta f_G g_G) + (1-\mu)(g_B + \beta f_B g_B)] - 2R_f$ and $B = \beta(1-\beta)(\mu g_G + (1-\mu)g_B)$.

We assume that $B' < 1$ to avoid perpetual growth.[33] In this case, $k_{t+1} - k_t < 0$ to the left and above of (16), and $S_{t+1} - S_t > 0$ to the left and above of (17).

By setting the restrictions below, we essentially focus on a particular configuration of the phase diagram, one that allows for the possibility of equilibrium paths that exhibit slow expansions and sudden drops.

Assumption 7. Suppose that $\mu k_s^G > \bar{k}$ and $B' < 1$. In that case, the solution to (14) at $k = \bar{k}$ is given by $S_1 = \frac{2\bar{k}}{k_s^G}$, and the solution to (16) at $k = \bar{k}$ is given by $S_3 = \frac{2\bar{k}}{A' + B'\bar{k}}$. Denote by S_2 and S_4 the respective solutions to (15) and (17) at $k = \bar{k}$. Suppose that $S_4 < S_1 < \min\{S_2, S_3\}$.

Assuming that $S_1 < S_3$ is equivalent to the parametric restriction we made in Assumption 5. Assuming $S_4 < S_1$ ensures that the equilibrium path does not converge to a steady state with separating contracts, while $S_1 < S_2$ rules out the convergence to a steady state under pooling.

We find the parametric example satisfying Assumption 7 that gives rise to the cyclical dynamic equilibrium exhibiting sudden drops and slow recoveries. In fact, we also chose the parameters to illustrate the possibility of non-monotone expansions.

Example 3. (Fig. 7) *Cyclical Dynamics with Endogenous Savings. Suppose that* $R_f = 1$, $\beta = 0.5$, $\mu = 0.2$, $f_G = 0.02$, $f_B = 0$, $g_G = 15.3$, $g_B = 3.1$, $\xi = 0.17$. *There exists a cyclical dynamic equilibrium for* $(S_0, k_0) = (0.3, 2)$ *exhibiting 5-period cycles that feature slow (non-monotone) recoveries and sudden drops.*

In the example given above, capital and supply of funds are low at the trough. Separating contracts emerge, and the strong positive selection of projects drives capital up for the next period. Productivity and net worth of the financed entrepreneurs are high, implying high savings for this group of agents. However, the low capital level also implies that the income of the unfinanced entrepreneurs ($(1-\beta)k$) is quite low, implying low savings for this group. In our example, this negative effect slightly dominates at this point of the cycle, resulting in less loanable funds, and therefore lower investment and slightly lower capital accumulation for the period after. From this point on, a full recovery takes place, moving savings and capital steadily upwards until capital is large enough to induce a pooling contract. At that point, the financing of type G entrepreneurs leads to a sudden drop in productivity and capital stock.

Because we construct the proposed equilibrium path using the appropriate contract determination and the appropriate evolution of capital and savings, to prove that the resulting paths comprise the dynamic equilibrium, it remains to check that the implied prices satisfy the conditions of Assumption 3. These can be easily checked using the equilibrium price definitions and the obtained path of capital accumulation $\{k_t^*\}$.

In the example economy, capital, savings, investment, loanable funds, output, cash flows and net worth move together. These variables are procyclical. Default rates are at their highest following

[33] This is slightly stronger than Assumption 6, as it ensures no perpetual growth with all entrepreneurs entering production under pooling, not only a fraction μ of them.

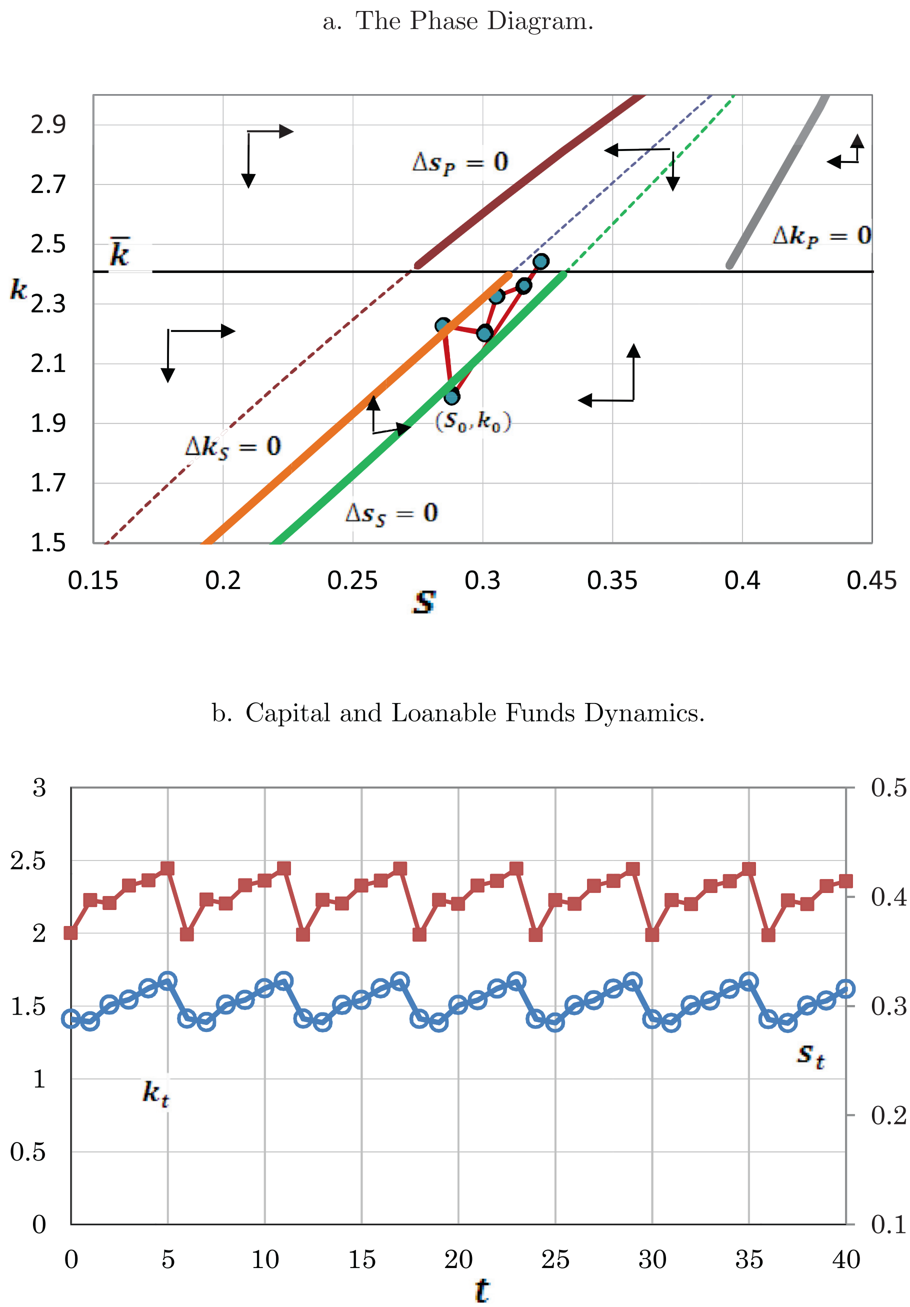

Fig. 7. Phase diagram and cyclical dynamics for the model with the endogenous supply of funds. Notes: this figure illustrates the phase diagram and cyclical dynamics for the model with endogenous savings, for the parameters chosen in Example 3 of Section 5. We observe that cyclical dynamic equilibria can exhibit sudden drops and slow (possibly non-monotone) recoveries.

the economic peaks. Hence, the extended model generates predictions that are qualitatively consistent with the relevant empirical evidence.

6. Conclusions

While the previous work modeling financial market frictions as a source of macroeconomic instability focused primarily on the role of moral hazard, we emphasize the importance of adverse selection and producer composition effects.[34] Endogenous cycles ensue via the empirically plausible dynamics of screening standards. In fact, our model provides new insight into why screening standards vary systematically over the business cycle in both credit and equity markets, with laxer terms applied in economic peaks and tighter terms applied in economic troughs. The model dynamics is qualitatively consistent with the dynamics of net worth, cash flows, investment, average loan size, default rates and entering producer quality over the business cycle.

The main insight derives from the dynamics of cash flows enjoyed by the entrepreneurs. As cash flow rises, tighter investment restrictions are needed in order to exclude the unproductive types. This high signaling cost makes separating contracts unattractive to productive entrepreneurs. As a result, suboptimal pooling con-

[34] Moral hazard is also present in our model due to the presence of default choice, but it plays no role in cyclical dynamics as it is always resolved in the same way.

tracts emerge in financial markets. Screening is abandoned and the unproductive entrepreneurs are financed alongside the productive types. Although aggregate investment rises as less resources are allocated to screening, the composition of entering entrepreneurs deteriorates, sending the economy into a recession. Consequently, weak economic conditions and low cash flow lead to tighter lending terms, the strong composition effect generating economic recovery.

For future research, we suggest building these features into a richer general equilibrium environment designed to study business cycles quantitatively. The objective would be to perform a quantitative assessment of the role played by adverse selection in financial markets in slowing down or even reverting positive trends in aggregate productivity. Indeed, since the 2008 financial crisis, financial shocks – modeled in a reduced form – have taken a prominent role in business cycle literature. But the underlying financial frictions, which seem critical for understanding recession triggers, remain far from well understood. Our explanation is that higher cash flow increases the cost of signaling and leads to abandoned screening, allowing the lemons to enter the mix of producers and trigger a recession. Note that one would generally expect the mix of potential applicants to deteriorate as the expansions persist, as the best ones are typically the first ones to lock into profitable opportunities, but this change would not necessarily imply a resource misallocation, not to mention financing of the negative present value projects. We emphasize an additional effect on the mix of producers appearing due to the presence of adverse selection.

Acknowledgments

We are grateful to Gary Biglaiser, Philip Bond, Hugo Hopenhayn, Larry Jones, Lutz Hendricks, Kiminori Matsuyama, Roger Myerson, Toshihiko Mukoyama, Peter Norman, Adriano Rampini, Balazs Szentes and to seminar participants at the University of British Columbia, Brown University, University of Chicago, Universidad de Chile, Clemson University, Duke University, International Monetary Fund, University of Minnesota, University of North Carolina, University of Oregon, and University of Washington.

Appendix. More general contracts

Consider the financial market equilibrium obtained by the more general contract menu of the form: $\{(1+\varepsilon_i, 1-\varepsilon_i, \delta_i, R_i)\}_{i=G,B}$ where $\varepsilon_i \in (-1, 1)$. As in the paper, we define the financial equilibrium, for given ρ and w , by contracts $\{(1+\varepsilon_i, 1-\varepsilon_i, \delta_i, R_i)\}_{i=G,B}$ such that (1) they induce self-selection and participation by entrepreneurs and participation by banks, (2) banks make zero profits, (3) $(1+\varepsilon_G, 1-\varepsilon_G, \delta_G, R_G)$ is the best loan contract for type G among the contract menu $\{(1+\varepsilon_i, 1-\varepsilon_i, \delta_i, R_i)\}_{i=G,B}$ satisfying (1) and (2).

Claim. *Our claim is that the equilibrium contract menu* $\{(1+\varepsilon_i, 1-\varepsilon_i, \delta_i, R_i)\}_{i=G,B}$ *will feature* $\varepsilon_i = 0$.

Proof. It is immediate that backloaded contracts with $\varepsilon_i < 0$ will not be offered in equilibrium because they preclude production. It remains to rule out $\varepsilon_i > 0$ for both i.

Consider the equilibrium contract menu of the form $\{(1+\varepsilon_i, 1-\varepsilon_i, \delta_i, R_i)\}_{i=G,B}$ where $\varepsilon_i > 0$ for both i. Self-selection implies that each type i prefers contract $(1+\varepsilon_i, 1-\varepsilon_i, \delta_i, R_i)$.

Case 1. Suppose the equilibrium contract menu is such that type B chooses not to enter, i.e. he cannot afford the early payment under any of the two contracts: $\delta_i > w + \rho f_B + \varepsilon_B$. Consider the following alternative menu: $(1, 1, \delta_G - \varepsilon_G, R_G)$ for type G and $(1, 1, \delta_B - \varepsilon_B, R_B)$ for type B. Under this menu, entry is still unaffordable for type B, but type G is strictly better off. Banks' participation and profits are unchanged. This implies that the original contract menu fails to satisfy part (3) of the definition, a contradiction.

Case 2. Suppose the equilibrium contract menu is such that type B chooses to enter. Self-selection then implies that $\delta_B \leq \delta_G$, or else type B would choose the contract offered to G. This is because type B defaults and hence only cares about having a low δ. Consider an alternative contract for type B given by $(1, 1, 0, \tilde{R}_B)$, which implies that B still enters and the banks' collections drop by $(1-\mu)\delta_B$. Consider an alternative contract for type G given by $(1, 1, \delta_G - \delta_B, R_G + \frac{1-\mu}{\mu}\delta_B)$. Note that in both contracts, the early repayment declined by the same amount and hence self-selection for B still holds. We set $\tilde{R}_B$ sufficiently high to make sure G does not want to switch. Type G is strictly better off under the new contract because the gain from the lower repayment more than compensates for the loss due to the higher interest rate. This is due to Assumption 4(a) which guarantees that δ is costly for type G under pooling. Banks' participation and profits are unchanged. Again, this implies that the original contract menu fails to satisfy part (3) of the definition, a contradiction. □

Proof of Lemma 1

Since B's payoff decreases in δ_B and is independent of R_B, self-selection for B implies $\delta^B \leq \delta^G$. Since B enters to default, bank participation requires that collections from type G must compensate the bank for the losses associated with lending to type B. Suppose $\delta^B < \delta^G$. Then a pooling contract (δ^B, R^G) can be offered. Type G strictly prefers it to (δ^G, R^G) because it offers a strictly lower early payment. Self-selection is satisfied, and banks make zero profits because collections from each type are unchanged. Therefore, we arrive at a contradiction with $\delta^B < \delta^G$. Hence, it must be the case that $\delta^B = \delta^G$.

It follows that it must also be the case that $R^B \geq R^G$, or else type G would never select his contract. Therefore, if B is financed, he is financed under a pooling contract.

Proof of Lemma 2

1. Fix an arbitrary $\delta < \tilde{\delta}(w, \rho)$. We first show that $R(w, \rho, \delta) > R_f$, i.e. that the pooling contract implies a cross-subsidy. From the pooling interest rate (5), we see that it decreases in δ. Therefore it suffices to show that the result holds for the smallest possible interest rate under pooling, i.e. $R\left(w, \rho, \tilde{\delta}\right)$. The difference $R\left(w, \rho, \tilde{\delta}\right) - R_f > 0$ simplifies to $2R_f - \tilde{\delta} > 0$, which holds because the early loan payment (in the case of separation) cannot exceed the total payment.

Next we show that the pooling contract (δ, $R(w, \rho, \delta)$) induces G to enter, i.e.

$$\rho g_G[\rho f_G + 1 - \delta + w] - 2R + \delta \geq w.$$

Using the zero profit condition, the above inequality can be rewritten as

$$\rho g_G[\rho f_G + 1 - \delta + w] - \frac{2R_f}{\mu} + \frac{1}{\mu}\delta \geq w. \tag{18}$$

The left hand side illustrates that δ reduces the amount of investment, but also lowers the required cross-subsidy. Assumption 4(a) ensures that the direct negative effect of δ dominates. It follows that the critical case is $\tilde{\delta}(w, \rho)$. Rewriting (18) for $\tilde{\delta}(w, \rho) = w + \rho f_B$ gives

$$\rho g_G[\rho f_G + 1 - \rho f_B] - \frac{2R_f}{\mu} + \frac{1}{\mu}(w + \rho f_B) \geq w$$

which, in light of $f_G > f_B$ and $\mu \in (0, 1)$, is implied by

$$\mu \rho g_G + \rho f_B \geq 2R_f,$$

i.e. Assumption 4(b).

2. We finally show that the separating menu offering $(\tilde{\delta}(w,\rho), R_f)$ to G and non-participation to type B induces G to enter. By the first result of the present lemma and by continuity, we know that type G decides to enter and repay when offered a contract $\left(\tilde{\delta}(w,\rho), R\left(w,\rho,\tilde{\delta}(w,\rho)\right)\right)$. Replacing the interest rate by $R_f < R\left(w,\rho,\tilde{\delta}(w,\rho)\right)$, also known by the first result, makes the repayment option more attractive without changing the non-entry options. It follows that type G enters.

Proof of Proposition 2

Consider $\{k_t^*\}_{t\geq 0}$ defined inductively as $k_0^* = k_0$, and for a given k_t^*, define k_{t+1}^* as in (11). Let $k_{\min}^* := \inf_t k_t^*$ and $k_{\max}^* := \sup_t k_t^*$. Define a sequence of beliefs $\{k_t'^*\}_{t\geq 0}$ by setting $k_t'^* = k_{t+1}^*$, and price sequences by $w_t^* = (1-\beta)k_t^*$ and $\rho_t^* = \beta$. Given these prices, define (R_t^*, δ_t^*) as in Proposition 1. It remains to check that Assumptions 1–6 are satisfied for ρ_t^* and w_t^* for all t. Noting that $w(k)$ is increasing in k, the hypothesis guarantees exactly this as long as $k_{\max}^* \leq k_s$. We prove this by showing inductively that $k_t^* < k_s$ for all t.

We have $k_0^* \leq k_s$ by hypothesis. Also, $k_{t+1}^* \leq \max\{k_s, k_p(k_t^*)\} \leq \max\{k_s, k_p(k_s)\} \leq \max\{k_s, k_s\} = k_s$ where the second inequality is due to the induction hypothesis and the fact that $k_p(k)$ is increasing, and the third inequality is implied by Assumptions 5 and 6.

Now, to see the existence of cycles, consider $k_0 < \bar{k}$ (the other case is analogous). Then $k_1 = k_s > \bar{k}$. It remains to show that $\exists N$ such that $k_p^N(k_1) < \bar{k}$. Since $k_p(k)$ is a linear function, we can write $k_p(k) = a + bk$, with $b < 1$ by Assumption 6. This, together with Assumption 5, implies that $k_1 - k_p(k) > 0$. Then $k_n - k_{n+1} = b^n[(1-b)k_1 - a] = b^n[k_1 - k_p(k_1)] \to 0^+$. Since by hypothesis $\bar{k} - k_p\left(\bar{k}\right) > 0$, $\exists N$ such that $k_p^N(k_1) < \bar{k}$. Then $k_{N+1} = k_s$ and the cycle is repeated.

Proof of Corollary 1

We know that $k_0 = k_s > \bar{k}$ and therefore $k_1 = k_p(k_s)$. For all $m \leq n$ we have then $k_m = k_p^{(m)}(k_s)$ (since by hypothesis $k_p^{(m-1)}(k_s) > \bar{k}$). Since, also by hypothesis, $k_m > k_s$, we have that $k_{m+1} = k_s$ and the result follows.

References

Asea, P.K., Blomberg, B., 1998. Lending cycles. J. Econom. 83 (1–2), 89–128.
Azariadis, C., Smith, B., 1998. Financial intermediation and regime switching in business cycles. Am. Econ. Rev. 88 (3), 516–536.
Berger, A.N., Udell, G.F., 2004. The institutional memory hypothesis and the procyclicality of bank lending behavior. J. Financ. Intermed. 13 (4), 458–495.
Bernanke, B., Gertler, M., 1989. Agency costs, net worth, and business fluctuations. Am. Econ. Rev. 79 (1), 14–31.
Bernanke, B., Gertler, M., Gilchrist, S., 1999. The financial accelerator in a quantitative business cycle framework. In: Taylor, J.B., Woodford, M. (Eds.), Handbook of Macroeconomics, 1, pp. 1341–1393.
Bester, H., 1985. Screening vs. rationing in credit markets with imperfect information. Am. Econ. Rev. 75 (4), 850–855.
Boissay, F., Collard, F., Smets, F., 2013. Booms and Systemic Banking Crises. European Central Bank. Working paper series 1514.
Caballero, R., Hammour, M., 2005. The cost of recessions revisited: a reverse-liquidationist view. Rev. Econ. Stud. 313–341.
Carlstrom, C., Fuerst, T., 1997. Agency costs, net worth, and business fluctuations: a computable general equilibrium analysis. Am. Econ. Rev. 87 (5), 893–910.
Cordoba, J.C., Ripoll, M., 2004. Credit cycles redux. Int. Econ. Rev. 45, 1011–1046.
De Meza, D., Webb, D.C., 1987. Too much investment: a problem of asymmetric information. Q. J. Econ. 102 (2), 281–292.
Dell'ariccia, G., Marquez, R., 2006. Lending booms and lending standards. J. Financ. 61 (5), 2511–2546.
Favara, G., 2012. Agency problems and endogenous investment fluctuations. Rev. Financ. Stud. doi:10.1093/rfs/hhs009. First published online March 16.
Figueroa, N., Leukhina, O., 2015. Lending standards and their impact on productivity. J. Econ. Dyn. Control 59, 1–21.
Gertler, M., Kiyotaki, N., 2010. Financial intermediation and credit policy in business cycle analysis. The Handbook of Monetary Economics.
Gompers, P., Kovner, A., Lerner, J., Scharfstein, D., 2008. Venture capital investment cycles: the impact of public markets. J. Financ. Econ. 87 (1), 1–23.
Greenwald, B.C., Stiglitz, J.E., 1993. Financial market imperfections and business cycles. Q. J. Econ. 108 (1), 77–114.
Hellwig, M., 1987. Some recent developments in the theory of competition in markets with adverse selection. Eur. Econ. Rev. 31 (1–2), 319–325.
Jermann, U., Quadrini, V., 2012. Macroeconomic effects of financial shocks. Am. Econ. Rev. 102 (1), 2012.
Khan, A., Thomas, J.K., 2013. Credit shocks and aggregate fluctuations in an economy with production heterogeneity. J. Political Econ. 121 (6), 1055–1107.
Kindleberger, C.P., 1996. Manias, Panics, and Crashes: A History of Financial Crises, Third ed. John Wiley & Sons, Inc, New York.
Kiyotaki, N., Moore, J., 1997. Credit cycles. J. Political Econ. 105 (2), 211–248.
Lee, Y., Mukoyama, T., 2012. Entry, Exit and Plant-level Dynamics Over the Business Cycle. Federal Reserve Bank of Cleveland. Working Manuscript.
Lown, C., Morgan, D.P., 2006. The credit cycle and the business cycle: new findings using the loan officer opinion survey. J. Money Credit Bank. 38 (6), 1575–1597.
Martin, A., 2008. Endogenous Credit Cycles. Universitat Pompeu Fabra. Working Manuscript.
Matsuyama, K., 2013. The good, the bad, and the ugly: an inquiry into the causes and nature of credit cycles. Theor. Econ. 8, 623–651.
Minsky, H., 1992. The Financial Instability Hypothesis. Jerome Levy Economics Institute, Annandale on Hudson, NY. Working Paper 74.
Myerson, R.B., 2012. A model of moral-hazard credit cycles. J. Political Econ. 120 (5), 847–878.
Rajan, R.G., 1994. Why bank credit policies fluctuate: a theory and some evidence. Q. J. Econ. 109 (2), 399–441.
Rampini, A., 2004. Entrepreneurial activity, risk, and the business cycle. J. Monet. Econ. 51 (3), 555–573.
Reichlin, P., Siconolfi, P., 2004. Optimal debt contracts and moral hazard along the business cycle. Econ. Theory 24 (1), 75–109.
Rothschild, M., Stiglitz, J.E., 1976. Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. Q. J. Econ. 90 (4), 630–649.
Ruckes, M., 2004. Bank competition and credit standards. Rev. Financ. Stud. 17 (4), 1073–1102.
Stiglitz, J.E., Weiss, A., 1981. Credit rationing in markets with imperfect information. Am. Econ. Rev. 71, 393–410.
Suarez, J., Sussman, O., 1997. Endogenous cycles in a stiglitz-weiss economy. J. Econ. Theory 76, 47–71.
Williamson, S., 1987. Financial intermediation, business failures, and real business cycles. J. Political Econ. 95, 1196–1216.