Technology Adoption, Market Structure, and the Cost of Bank Intermediation*

Gianni De Nicolò Andrea F. Presbitero Alessandro Rebucci Gang Zhang

Monday 22nd March, 2021

Preliminary. Comments Welcome.

Abstract

This paper studies the high and persistent U.S. cost of financial intermediation (CFI) documented by Philippon (2015) and its inverted U-shape behavior since the mid-1960s. We build a novel model of endogenous growth and bank intermediation and introduce imperfect bank competition, bank IT adoption and bank entry, and an occupational choice that determines the relative size of the labor force and the economy's average level of managerial ability. The interplay between verification costs, market structure, and occupational choice delivers implications for the CFI which are qualitatively consistent with the stylized facts of the U.S. economy. We find that the banking sector structure is the main determinant of the long-run level of the CFI. We also show that the U.S. productivity growth slowdown from the mid-1960s to the mid-1980s is a major driver of the simultaneous increase in the CFI and the number of banks during this period and their subsequent decline.

Keywords: Banking, Cost of Finance, Financial Intermediation, Imperfect Competition, Productivity slowdown, Technology Adoption.

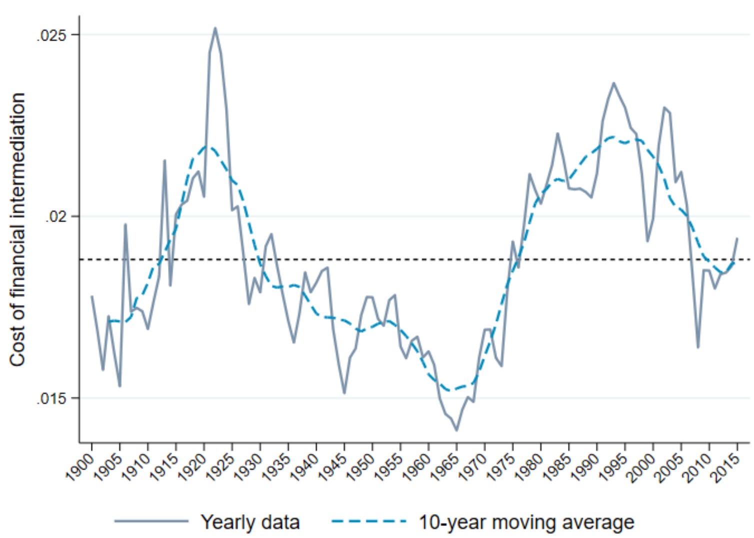
JEL Classification: E44; G21; G38; O33

*Gianni De Nicolo', JHU Carey Business School, email: gdenico1@jhu.edu; Andrea F. Presbitero, JHU School of Advanced International Studies and CEPR, email: apresbitero@jhu.edu; Alessandro Rebucci, JHU Carey Business School, CEPR and NBER, email: arebucci@jhu.edu (corresponding author); Gang Zhang, Cheung Kong GSB, email: gzhang@ckgsb.edu.cn. We thank Laurence Ball, Nicolas Crouzet, Olivier Jeanne, Thomas Philippon, Qi Sun, Xiaodong Zhu, and participants to the 2nd IMF Macro-Financial Research Conference, the Fall 2019 Midwest Macroeconomics Meetings, and seminars at CKGSB, Fudan, JHU, STJU, SUFE, Tsinghua, and ZJU for useful comments and discussions. We thank Huy Nguyen and Huajun Meng for excellent research assistance. Rebucci is grateful the the Johns Hopkins University Catalyst Award Program for financial support. Zhang gratefully acknowledges financial support from the ASEAN Business Research Initiative.

1 Introduction

Understanding the relationship between an economy's long-run growth potential and the cost of financial intermediation is critical in assessing the extent to which the benefits of finance are commensurate to its costs. In the United States, the cost of financial intermediation, defined as the ratio of value added of the financial industry over total intermediated assets, henceforth denoted CFI, has averaged about 2 percent since the late 1800s (Philippon, 2015, 2016).[1] As Figure 1 shows, however, the CFI has also fluctuated significantly at a low frequency of 10 years or lower. For example, it fell steadily from a peak of more than 2.5 percent in the early-1920s to a trough well below 1.5% in the mod-1960s. It then climbed again well above 2% early 1990s, to decline more steeply in the 2000s.

Figure 1: U.S. Cost of Financial Intermediation (1900-2017)

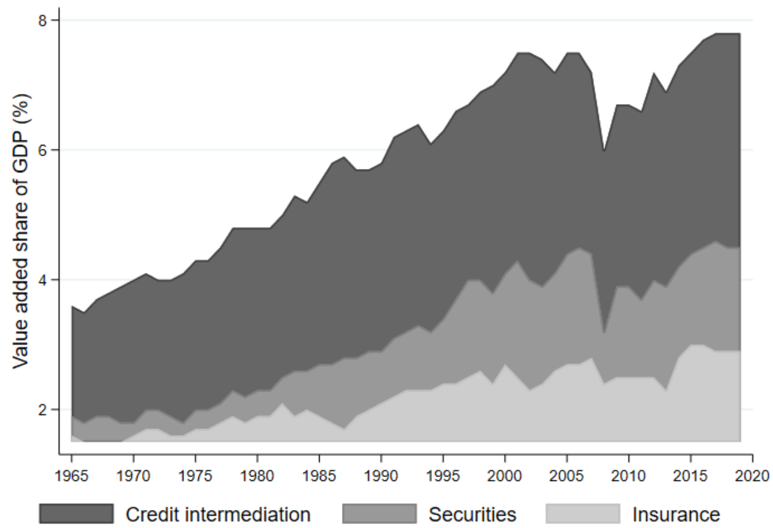


Notes: The figure shows the cost of financial intermediation in the United States. The solid line is an updated version of Philippon (2015) estimate. The dashed line is a 10-year moving average. The dotted horizontal line is long-term average computed over the period 1900-2017. Source: Authors' calculation based on Philippon (2015). Details on the methodology and a comparison with the original data are in the Data Appendix.

In this paper we develop an endogenous growth model with bank intermediation whose balance growth path matches the long run value of the CFI. We then zoom in the the last 50 years of history and provide a model-based account of the CFI's inverted U-shape behavior since 1965 in Figure 1, investigating the mechanisms through which economy-wide productivity and

[1] Bazot (2017) reports similar findings for Europe and the rest of the world, respectively.

Figure 2: THE U.S. FINANCE INDUSTRY (1965-2019)



NOTES: The figure reproduces and updates the evidence in Greenwood and Scharfstein (2013), plotting the share of value added of credit intermediation, securities and insurance in GDP as in Figure 1 of Greenwood and Scharfstein (2013). Source: Flow of Funds Data from Data from the Bureau of Economic Analysis.

other structural changes in the U.S. economy can drive low frequency fluctuations in the CFI.

We focus on the banking sector since credit intermediation continues to represents the largest segment of the U.S. financial sector. As Figure 2 illustrates, in fact, the banking sector value added accounts for about 50% of the total in the financial sector. Specifically, we study the determinants of the long-run value of the cost of bank intermediation and its evolution since the mid-1960s, when the CFI troughs and starts to rise. Indeed, Appendix B shows that the cost of *bank* intermediation (henceforth CBI), defined as the ratio of value added over intermediated assets in banking, follows a similar pattern as the CFI since 1976 when the two series overlap.

The model introduces Cournot bank loan competition, bank information technology (IT) adoption, and an endogenous occupation choice in a standard costly state verification setup (e.g., Townsend, 1979; Williamson, 1986; Bernanke, Gertler and Gilchrist, 1999; Greenwood, Sanchez and Wang, 2010) with bank entry. Agents in the economy choose to be workers or entrepreneurs. This occupation choice determines the relative size of the labor force and the average level of managerial ability in the economy, which the main driver of endogenous growth in our model. The model also features a bank IT adoption decision that determines lending rates in a general equilibrium framework in which banks compete imperfectly in the loan market facing costly state verification. Third and finally, the model also allows for bank entry *à la* Melitz (2003) driven by expected bank profit and a wage-based reservation value. The model yields a rich two-way interaction between finance and growth in which small changes in the CFI lead to large changes in long-run growth. The model's application to the analysis of the long-term value of the CBI and its low frequency changes shows that market structure is more important than information asymmetry in explaining the steady state value of the CBI. A counterfactual simulation of the model also illustrates that the U.S. productivity slowdown can go a long way in accounting for the inverted U-shape of the CFI and the number of banks in the system since the mid-1960s.

In the model, a set of heterogeneous agents sort themselves into entrepreneurs or workers. Entrepreneurs hire workers to produce capital goods, while workers supply one unit of labor inelastically. Entrepreneurs borrow from banks to pay wages upfront. The firms' marginal cost (profit), therefore, is increasing (decreasing) in the lending rate. Banks know the entrepreneur type, but can verify the ex ante uncertain realization of the enterpreneur's ability (or the success of their project) only exerting costly effort and adopting IT equipment. IT equipment adoption allows banks to chose the individual level of operating efficiency. If the entrepreneur's project fails, he/she defaults on the bank loan.

In the loan market, banks compete *à la* Cournot-Nash given a perfectly inelastic supply of deposits at a given deposit rate, as in Greenwood *et al.* (2010). The potential bank entrants' reservation value takes the form of labor and represents the fixed cost of setting up a bank, including infrastructure costs for branches, ATMs, IT and legal architecture. A representative household, that owns a final good firm produced with capital goods as input, closes the model. The final good is consumed by the household, entrepreneurs, workers, and bankers. In order to keep the model tractable, we abstract from aggregate uncertainty and from the household's consumption-saving decision.

Aggregate technology and banking efficiency evolve endogenously. Following Melitz (2003), we assume that both aggregate processes are defined as the average across all individuals of the individual characteristics. In order to obtain a well-behaved balance growth path dynamics, we normalize aggregate banking sector productivity relative to the non-financial one. In addition, we also adjust this average for the relative share of workers and entrepreneurs. The adjustment is interpretable as industry organizational capital–e.g., Atkeson and Kehoe (2005) and McGrattan and Prescott (2005).

The contemporaneous technological progress in the non-financial sector influences the growth of the aggregate banking efficiency, and vice-versa. Consider, for example, a rise in aggregate non-financial sector productivity. It boosts capital good output, entrepreneurs' profits, labor and credit demand, and lowers the capital good price. Higher entrepreneurs' profits implies higher banks' profits. Higher labor demand drives up the wage. The capital good price decline induces banks to adopt more IT equipment, improving their individual operating efficiency and allowing them to lower the loan rate.

The productivity increase raises the wage, credit demand, and the firm profit. The bank profit also rises. The wage directly increase bank reservation values. However, the interest rate decline partially offsets the increases in banks' total profits. As a result, the reservation value increase is larger than the increase in bank profit, which leads fewer banks to enter.

Falling lending rates in response to an aggregate productivity increase feed back to the real economy in two ways. First, cheaper financing induces entrepreneurs to further expand their production. Second, because of lower lending rates, the marginal cost of labor and the firm profit increase less than the rise in the wage. This leads more agents to choose to be workers and raises the threshold for the occupation choices. As a result, aggregate end-of-period productivity, defined as the average of managerial ability across all entrepreneurs, is higher.

Higher individual banking efficiency triggered by IT adoption in response to an aggregate productivity increase, all else equal, leads to higher average banking sector efficiency. However, lending to a more productive pool of entrepreneurs, banks accumulate less organizational capital, and aggregate banking efficiency could rise less than the individual ones or decline.

Similar dynamics are at play in the transmission of an exogenous change in aggregate banking sector efficiency. An aggregate banking efficiency increase leads to lower lending rates charged by banks, higher capital good output and wages. Lower rates also imply fewer banks entering the market since the bank profit increases less than the wage. Entrepreneurs's profits go up, but less than the wage that needs to be prepaid, thanks to the lower lending rates. More agents choose to become workers, the threshold value for the occupation choice increases, and the average managerial ability in the economy rises. As a result aggregate non-financial sector productivity is also higher. More supply of capital goods pushes down their price, inducing banks to adopt more IT equipment, improve their individual efficiency, and lower lending rates further.

In the model, the CBI is defined as the ratio of the total profit in the banking industry over total loans borrowed by entrepreneurs. The CBI is affected by both the loan market structure and the interest rate, both of which reflect the financial friction due to the costly state verification and imperfect competition. In particular, more banks lead to more competition, reducing the banks' pricing power, which result in a lower CBI. As banks become more efficient at verifying the states, they charge a lower lending rate, which implies a lower CBI. Furthermore, a lower rate induces fewer banks to enter. Fewer banks mean lower competition and more market power. Note that the decline in the number of banks increases the pricing power as well as the market share of banks. The rise in the pricing power only partially offsets the decline in the CBI, as the rise in market share does not affect the CBI at the industry level.

Under plausible parameter restrictions, the model economy admits a unique and stable balance growth path in which the threshold value for the occupation choice, the number of banks, the interest rate charged by banks, and the CBI are constant, while the economy grows and the price of capital goods declines steadily.

We calibrate the balance growth path of the economy to basic stylized facts of the U.S. long-run growth. We then counterfactually remove the information asymmetry by assuming that banks can always verify the entrepreneurs' true states. We find that the CBI declines by only 3.7%, from 1.88% to 1.81%, while the interest rate spread declines from 8% to 1.77%. When we remove costly verification, the decline in the CBI is partially offset by the drop in the number of banks. Importantly, however, the long-run growth rate of the labor productivity increases by almost 30%, from 2.1% to 2.7%. The associated decline in the interest rate lowers the effective marginal cost of labor that has to be paid on borrowed funds and stimulates output growth. As a result, labor productivity increases. We interpret this result as suggesting that imperfect competition is more important than costly state verification in driving the long-run value of the CBI, as in our model there is no value added from labor in banking and in the data this component of the CBI is relatively stable over time in the aggregate. This highlights the importance of financial development for economy growth, as in King and Levine (1993); Jayaratne and Strahan (1996); Rajan and Zingales (1998); Levine, Loayza and Beck (2000); Greenwood and Scharfstein (2013), among others.

We then consider the potential impact on the CBI and the banking industry of the U.S. productivity growth slowdown starting in the mid-1960s and an increase in the bank funding costs associated with the removal of Regulation Q and the Volcker disinflation in the early 1980s. The model-implied path of the CBI and the number of banks is qualitatively consistent with that in the data. We find that the model can generate the inverted U-shape in the CBI, the number of banks, and the degree of market concentration that we document in the data. The model also tracks relatively well the path of banks' intangible assets and the loan-deposit interest spread. The model can match the magnitude of the changes in market concentration and the number of banks. However, it underestimates the magnitude of the changes in the CBI and intangible assets.

Our model integrates elements from three strands of literature. First, the paper relates to the growing literature on market structure, regulation and competition in banking (see, for example, Corbae and Levine, 2020). Second, the paper belongs to the literature on technology adoption and growth (e.g., Aghion, Howitt and Levine, 2018; Aghion, Bergeaud, Boppart, Klenow and Li, 2019). Finally, the paper is connected to contributions that introduce banking and financial frictions in standard growth frameworks (see, e.g., Gersbach, Rochet and Scheffel, 2019; Greenwood, Sanchez and Wang, 2010). As far as we are aware, this is the first paper to attempt to provide a model account of the long-run determinants of the CFI and its low frequency changes.

The reminder of the paper is organized as follows. Section 2 sets up the model. Section 3 derives the equilibrium and its properties. Section 4 describes the key properties of the model's solution and reports the main comparative static results on the CBI. Section 5 derives the balance growth path of the economy. Section 6 calibrates the balanced growth path of the economy to match the long-run value of the CFI. Section 7 presents the counterfactuals and discusses the model's ability to generate an inverted U-shape CFI as we see in the data from the mid-1960s to the 2010s. Section 8 concludes. The Appendix contains a description of the data (Section A), a comparison between the cost of intermediation in finance and in banking (Section B), and details on the the model solution as well as proofs of all propositions and lemmas (Section C).

2 A model of growth with bank intermediation

In this section we set up an endogenous growth model with financial intermediation to explain the CFI and evaluate its impact on the U.S. economy. The model features endogenous growth in both economy-wide and banking sector productivity, costly state verification, imperfect competition a la Cournot on the lending side, bank IT adoption, and an occupation choice driven by heterogeneous productivity.

The agents in the economy are producers of final goods, bankers, and agents who choose to become either workers or entrepreneurs producing capital goods used in final good production. For simplicity, we assume a perfectly inelastic supply of deposits at a given exogenous deposit rate.[2]

2.1 Final good producers

There is a set of risk-neutral agents of unit mass who produce the final good, Y_t, using a capital good as an input with the following production function:

$$Y_t = K_t^{\alpha}, \tag{1}$$

where, omitting individual indexation of the variables for simplicity, K is the aggregate level of capital per worker. Capital depreciates fully every period, and the capital share $\alpha \in (0, 1)$. The final good is the numeraire of the economy, and its price is normalized to 1. Final good producers choose capital to maximize profits, taking the price of the capital good, p_t, as given:

$$\max_{K_t} K_t^{\alpha} - p_t K_t. \tag{2}$$

The final good firm's first order condition is

$$\alpha K_t^{\alpha-1} = p_t, \tag{3}$$

yielding the following capital demand:

$$K_t = \left(\frac{\alpha}{p_t}\right)^{\frac{1}{1-\alpha}}. \tag{4}$$

The final good firm's profit is given by:

$$\pi_t^F = (1-\alpha)\left(\frac{\alpha}{p_t}\right)^{\frac{\alpha}{1-\alpha}}. \tag{5}$$

[2] The main results of the paper hold in a setting in which the deposit rate is endogenous. All results not reported in the paper are available on request from the authors.

2.2 Entrepreneurs and workers

There is a continuum of risk-neutral individuals (or agents) with M unit mass, each of whom is assigned a managerial ability θ drawn from a distribution, on the domain $\Theta = \left(\underline{\theta}, \bar{\theta}\right)$, with a c.d.f. $\mathbf{F}(\theta)$. The parameter θ characterizes their type, either as a worker inelastically supplying one unit of labor, or becoming an entrepreneur operating a firm which produces capital goods. Individuals are reassigned their type every period.

2.2.1 Entrepreneurs

The type–θ agent who chooses to become an entrepreneur has access to the following stochastic technology for the production of the capital good, k_t, using labor as sole input:

$$k_t(\theta) = \begin{cases} \theta z_{t-1} \left(l_t(\theta)\right)^{\xi} & \textit{with probability } \eta \\ 0 & \textit{with probability } 1-\eta \end{cases} \tag{6}$$

where $l_t(\theta)$ is the labor employed by the entrepreneur, $\xi \in \left(0, \frac{2}{3}\right)$ is the labor share, and z_{t-1} is the lagged level of aggregate technology in the capital good producing sector whose evolution is spelled in Section 5. The source of idiosyncratic uncertainty is the realization of type-θ managerial ability. If the managerial ability θ materializes, the entrepreneur will produce $\theta z_{t-1} \left(l_t(\theta)\right)^{\xi}$ units of capital goods. If it does not, production is nil. The realization of the managerial type is private information for the entrepreneur giving rise to a moral hazard problem.

Entrepreneurs have no resources and must finance production with bank loans. By feasibility, banks will receive no payment if the entrepreneur's project fails as no output is produced. Before learning the realization of her managerial ability, the type–θ entrepreneur chooses labor $l_t(\theta)$, taking as given the relative price of capital, p_t, the wage w_t, and the gross loan rate $r_t^L(\theta)$, to maximize her expected profit given by

$$\max_{l_t(\theta)} \eta \left(p_t \theta z_{t-1} \left(l_t(\theta)\right)^{\xi} - r_t^L(\theta) w_t l_t(\theta)\right). \tag{7}$$

The first order condition is

$$\xi p_t \theta z_{t-1} \left(l_t(\theta)\right)^{\xi-1} = r_t^L(\theta) w_t. \tag{8}$$

Thus, the amount borrowed from banks is given by

$$b_t(\theta) = wl_t(\theta) = \left(\frac{\xi p_t \theta z_{t-1}}{r_t^L(\theta) w_t^{\xi}}\right)^{\frac{1}{1-\xi}}. \tag{9}$$

Given the relative price of capital, p_t, the wage w_t, and the gross loan rate r_t^L, the type–θ entrepreneur's expected profit is

$$\pi_t^E(\theta) = \eta\left(\frac{1}{\xi} - 1\right)\left(\frac{\xi p_t \theta z_{t-1}}{\left(r_t^L(\theta) w_t\right)^{\xi}}\right)^{\frac{1}{1-\xi}}. \tag{10}$$

Entrepreneurs' profits are higher with a higher price of capital goods or aggregate technology level and they decline as the wage or the borrowing cost increases. Note finally that, if $\theta > \theta'$, the type–θ agent has a comparative advantage in producing capital goods relative to the type–θ', since her managerial skill level is higher.

2.2.2 Workers

The type–θ agent who chooses to become a worker inelastically provides one unit of labor, and her total income is the wage w_t. Also, workers will be paid by entrepreneurs with borrowed funds from banks regardless of whether the entrepreneurs realize their managerial ability or not.

2.2.3 Occupation choice

Agents need to choose between being an entrepreneur or a worker, whatever generates a higher income. If $\pi^E(\theta) > w_t$, then the type–θ agent chooses to become entrepreneur. Otherwise, she provides labor as a worker. Thus, a more productive agent is more likely to become an entrepreneur because, all else equal, Equation (10) implies that a more productive agent can make a higher income as an entrepreneur.

2.3 Bankers

There is an infinitely number of potential bankers, who are agents endowed with a reservation value that can be used to cover for the fixed costs of setting up a bank. Each banker establishes and runs a single bank to enter the loan market. Once entered, banks compete *a la* Cournot in the loan market. The deposit market is perfectly competitive, and banks have a perfectly inelastic supply of deposits at a given net real interest rate, r^D. Due to the information asymmetry on the borrower side, banks need to monitor entrepreneurs to manage credit risk. Monitoring entails exerting costly effort to verify the states reported by entrepreneurs, as well as adopting information technology (IT) equipment for verification. Therefore, each bank has two separate decisions to make. The first one is the choice of the level of IT investment to determine their level of operating efficiency and to facilitate monitoring. The second one is the optimal level of effort to exert and the parameters of the incentive-compatible contract to offer to each entrepreneur, taking the level bank-specific efficiency as given.

Bank efficiency and IT adoption decision To simplify aggregation, we assume that banks transform deposits into loans less than one–for–one. Specifically, the bank j needs to collect $\frac{1}{a_j}$ of dollar deposits to issue one dollar of loans, where $a_{jt} < 1$ is the operating level of efficiency of bank j, which is endogenously determined as described below. Denote with $b_j(\theta)$ the size of the loan of bank j to the type–θ entrepreneur. Then, the bank needs to collect $\frac{b_{jt}(\theta)}{a_{jt}}$ of deposits.

Banks must choose their level of operating efficiency by investing in IT equipment before they lend. The bank j selects its operating efficiency, a_j, by acquiring q_j unit of capital goods. The cost of this IT investment in units of final goods is $p_t q_j$. Then bankers achieve their own operational efficiency by combining aggregate bank efficiency from the previous period with capital (IT) goods, according to:

$$a_{jt} = \kappa a_{t-1} q_{jt}^{\frac{1}{2}} \tag{11}$$

where κ is a scaling factor, and a_{t-1} is the aggregate level bank of efficiency from the previous period. Thus, the bank's individual efficiency a_j is an increasing and concave function in IT equipment q_j. Note that once adopted, banks can use their IT for all future lending by affecting next period aggregate banking efficiency.

Bank IT adoption provides a channel through which the real economy and financial development in the banking system, as measured improvement in a_{t-1}, interact. On the one hand, a decline in the price of capital goods encourages banks to adopt more IT, which improves the bank-specific efficiency and future aggregate banking efficiency. On the other hand, a higher level of banking efficiency lowers the borrowing cost of entrepreneurs, and hence can affect the wage rate, future aggregate productivity and economic growth.

Bank monitoring technology Because the entrepreneurs' realization of their managerial ability is private information, all bankers need to exert efforts and verify their borrowers reported states. We assume that all bankers have access to the same bank monitoring technology, and the required effort per unit of deposit is e_{jt}. The output of the monitoring technology is a probability of detecting the entrepreneurs' realization of their true managerial ability, θ.[3] The banker j's probability of detecting the true state of the tyep–θ entrepreneur is given by the following increasing and concave function in e_j:

$$\mathbf{Pr}(e_{jt}(\theta)) = \frac{e_{jt}(\theta)}{e_{jt}(\theta) + \sigma}, \tag{12}$$

where $\sigma > 0$ is a scaling factor. The total volume of deposits needed by the bank j for lending the amount b_{jt} to the type–θ entrepreneur is given by the principal amount of the loan plus the resources to fund exerting efforts and is given by

$$\frac{b_{jt}(\theta)}{a_{jt}} + (1-\eta)e_{jt}(\theta)\frac{b_{jt}(\theta)}{a_{jt}}. \tag{13}$$

Note that bankers only exert efforts when the entrepreneur reports that she does not realize her managerial ability. This is because entrepreneurs have no incentive to misreport when they do not realize their managerial ability, and exerting effort is costly.

Bankers' reservation value Suppose that the bankers' reservation value is equivalent to Γ units of labor with market value $w_t\Gamma$. We can interpret this reservation value as if bankers were to invest their own stock of human capital to acquire the necessary infrastructure, including branches, ATMs, and fund other fixed costs. Bankers keep entering the loan market as long as the expected profit exceeds this value. This is akin to the free-entry condition in Melitz (2003).

3 The intertemporal symmetric equilibrium

The sequence of agents' decisions at each date t is as follows:

1. At the beginning of each periods, individual agents are assigned a type drawn from a distribution over Θ. Then agents choose to become either entrepreneurs or workers conditional

[3]Our specification of the monitoring technology is germane to specifications of R&D activities analyzed in the IO literature–see for instance Dasgupta and Stiglitz (1980), Vives(1990), and Marshall, and Parra(2019) for a more recent contribution.

on z_{t-1} and a_{t-1}. Bankers choose whether to invest in infrastructure and enter the loan market taking the deposit rate, r^D, as given.

2. Banks adopt q_j units of capital goods, selecting their level of efficiency, a_j.

3. Entrepreneurs demand funds from banks. Banks compete by offering incentive compatible loan contracts to entrepreneurs in a symmetric loan market Cournot-Nash equilibrium.

4. Entrepreneurs managerial ability is realized, and their project succeeds or fails. Banks exert efforts to verify the entrepreneurs' state. Once the state is verified, entrepreneurs pay banks.

5. Final good producers demand the capital good from entrepreneurs.

6. The competitive capital good and labor markets clear.

7. All agents consume the final good and the competitive final good market clears.

8. The aggregate technology at capital good sector and the aggregate banking efficiency is assembled.

Denote C_t^F, C_t^E, C_t^W, and C_t^B the aggregate consumption of the final good producers, the entrepreneurs, the workers, and the bankers, respectively. An intertemporal symmetric equilibrium is defined as follows.

Definition 1 (Intertemporal symmetric equilibrium). *Given $\{z_{t-1}, a_{t-1}\}_{t=1}^{\infty}$ and the cost of funds, r^D, an intertemporal symmetric equilibrium is a sequence of prices $\{r_t^L, p_t, w_t\}_{t=1}^{\infty}$, number of banks $\{n_t\}_{t=1}^{\infty}$, banks' optimal contracts, $\{b_t, e_t, r_t^L\}_{t=1}^{\infty}$, and capital goods chosen by banks $\{q_t\}_{t=1}^{\infty}$, thresholds for the occupation choices $\{\theta_t^*\}_{t=1}^{\infty}$, the labor demands by the type–θ entrepreneur $\{l_t(\theta)\}_{t=1}^{\infty}$, and aggregate quantities $\{Y_t, K_t, C_t^F, C_t^E, C_t^W, C_t^B\}_{t=1}^{\infty}$ such that:*

1. *Given p_t, w_t, r_t^L, n_t, and θ_t^*, the final good producer and entrepreneurs maximize their profit;*

2. *Given p_t, w_t, n_t, and θ_t^*, banks optimally invest in IT equipment and maximize their profits in a symmetric loan market Cournot-Nash equilibrium;*

3. *Given n_t and θ_t^*, the lending rate r_t^L clears the credit market, the relative price of capital goods p_t clears the capital goods market, the wage w_t clears the labor market, and the final good market clears at the normalized final good price of 1;*

4. *Given θ_t^*, the number of banks n_t is determined by the free entry condition;*

5. *The threshold level of managerial ability θ_t^* is such that the marginal agent is indifferent between being a worker and an entrepreneur, and agents with a higher managerial ability chooses to be an entrepreneurs, whereas those with a lower ability to be workers.*

The solution of this equilibrium is obtained in five steps. First, we characterize the loan market equilibrium given the capital price p_t, the wage w_t, and the number of banks n_t. Second, we determine the equilibrium in the capital good and labor markets. Third, we determine the number of banks using the free entry condition. Forth, the cut-off point separating agents into workers and entrepreneurs is determined. Finally, the equilibrium in the final good market is determined.

3.1 Loan market Cournot equilibrium

At the beginning of time t, entrepreneurs with ability θ receive contracts from all n_t banks. These contracts specify a loan amount and the penalty rate applied to the misreporting entrepreneur after production has taken place, denoted $r_t^{\theta,0}$. To characterize this contract, note that no entrepreneur would misreport $\theta \geq 0$ if $\theta = 0$. Therefore, the penalty rate applies only to entrepreneurs who report $\theta = 0$ when it is $\theta > 0$. Furthermore, banks would put no effort in verifying the state of an entrepreneur who reports a successful project as banks observe the entrepreneur type. Since the probability that an entrepreneur does not realize her ability is $1 - \eta$, a bank would expect to exert efforts with the same probability. We can now define a symmetric Cournot-Nash loan market equilibrium as follows.

Definition 2 (**Symmetric Loan Market Cournot-Nash Equilibrium**). *Given the number of banks n_t, the bank–specific efficiency a_{jt}, the price of capital goods p_t, and the cost of funds r^D, a symmetric Cournot-Nash equilibrium in the loan market for the type θ entrepreneur is a set of bank contracts $(r_{jt}^L(\theta), r_t^{\theta,0}(\theta), b_{jt}(\theta))$, and a level of monitoring efforts, $e_{jt}(\theta)$, such that, for all $j \in \{1, 2, ..., n_t\}$:*

1. *Given a_{jt} and b_{it} for all $i \neq j$, bank j chooses r_t^L, $r_{jt}^{\theta,0}$, b_{jt}, and e_{jt} to solve:*

$$\max_{b_{jt}, r_{jt}^L, r_{jt}^{\theta,0}, e_{jt}} \quad \eta r_{jt}^L(\theta) b_{jt}(\theta) - \frac{r^D b_{jt}(\theta)}{a_{jt}} - (1-\eta)\frac{e_{jt}(\theta) r^D b_{jt}(\theta)}{a_{jt}} \tag{14}$$

$$s.t. \quad r_{jt}^L(\theta) = \frac{\xi p_t \theta z_{t-1}}{w_t^{\xi}} \left(\sum_{i=1}^{n_t} b_{it}(\theta)\right)^{\xi-1} \tag{15}$$

$$r_{jt}^{\theta,0}(\theta) b_{jt}(\theta) \leq \frac{b_{jt}(\theta)}{\sum_{i=1}^{n_t} b_{it}(\theta)} p_t \theta z_{t-1} \left(\sum_{i=1}^{n_t} b_{it}(\theta)\right)^{\xi} \tag{16}$$

$$r_{jt}^L(\theta) b_{jt}(\theta) \geq \mathbf{Pr}\left(e_{jt}(\theta)\right) r_{jt}^{\theta,0}(\theta) b_{jt}(\theta) \tag{17}$$

2. *And*

$$b_{jt}(\theta) = b_t(\theta) \text{ for all } j \in \{1, 2, ..., n_t\}. \tag{18}$$

According to (14), bank's expected operating profits from the type–θ entrepreneur is equal to expected net interest income minus expected monitoring costs. Equation (15) is the inverse loan demand, which is derived from Equation (9) by setting $\sum_{j=1}^{n_t} b_{jt}(\theta) = w_t l_t(\theta)$. Equation (16) is the resource constraint on the bank's penalty rate, which can penalize entrepreneurs only up to what they make when they are found misreporting. Constraint (17) is the incentive-compatibility constraint on entrepreneurs, ensuring that the profit of truth-reporting is at least as large as the expected profit when misreporting, given that the payment in case of failure is zero. Equation (18) imposes symmetry on the equilibrium. The following proposition characterizes the solution of the equilibrium in the loan market for the type–θ entrepreneur.

Proposition 1 (**Loan Market Equilibrium Solution**)**.** *The solution of the symmetric Cournot equilibrium in the loan market for the type θ entrepreneur is given by:*

$$r_{jt}^{\theta,0}(\theta) = \frac{p_t \theta z_{t-1}}{w_t^{\xi}} \left(n_t b_{jt}(\theta)\right)^{\xi-1} \tag{19}$$

$$e_{jt}(\theta) = \frac{\xi \sigma}{1-\xi} \tag{20}$$

$$\eta r_{jt}^L(\theta) = \frac{(1+\tilde{\sigma}) r^D}{\left(1 - \frac{1-\xi}{n_t}\right) a_{jt}}. \quad \text{where} \quad \tilde{\sigma} = \frac{(1-\eta)\xi\sigma}{1-\xi} \tag{21}$$

$$b_{jt}(\theta) = \frac{1}{n_t} \left(\frac{\eta p_t \xi \theta z_{t-1}}{w_t^{\xi}} \frac{\left(1 - \frac{1-\xi}{n_t}\right) a_{jt}}{(1+\tilde{\sigma}) r^D} \right)^{\frac{1}{1-\xi}} \tag{22}$$

Proof: see Appendix C.1.

The resource constraint (16) binds, implying that banks will take all revenue if they find that entrepreneurs misreport the project's failure. Equation (20) implies that the effort exerted by banks is the same for all entrepreneurs, and increases with σ. Equation (21) is the equilibrium risk-adjusted lending rate, ηr_t^L, where η is probability of project success. In fact, the lower is the probability of success, the higher higher the chance of default. This rate is also the same across all entrepreneurs, which depends on three factors: the first is the optimal monitoring costs, $\frac{1+\tilde{\sigma}}{a_j}$. The second is imperfect competition, $1 - \frac{1-\xi}{n_t}$. The third one is the deposit rate. Hence, banks charge a lower rate the higher their level of operating efficiency, the more competitive the loan market is, and the lower the deposit rate. Thus, the risk-adjusted lending rate is determined by the interplay between bank funding cost, monitoring costs, imperfect competition, and the bank's efficiency. Note here that both monitoring costs and bank efficiency, indirectly depends on the structural characteristics of the economy, through ξ, and its aggregate behavior, through a_j. Equations (22) is the loan supply and it implies that banks offer larger loans to more productive entrepreneurs, who require more funds to produce capital goods in larger projects.

3.2 Occupation choice

As Equation (10) shows, the expected profits of an agent with a level of managerial ability θ who chooses to become an entrepreneur depend on (i) the capital good price; (ii) the wage rate; (iii) the individual managerial ability; as well as (iv) the interest rate charged by banks. Combining Equation (10) with (21), we obtain

$$\pi_t^E(\theta) = \eta \left(\frac{1}{\xi} - 1 \right) \left(\eta \xi p_t \theta z_{t-1} \left(\frac{\left(1 - \frac{1-\xi}{n_t}\right) a_{jt}}{(1+\tilde{\sigma})\, r^D w_t} \right)^{\xi} \right)^{\frac{1}{1-\xi}}. \tag{23}$$

Thus, for any $\theta \in \Theta$, if $\pi_t^E(\theta) > w_t$, the type–θ agent chooses to be an entrepreneur; otherwise, she choose to be a worker. Henceforth, we denote the subset of agents who choose to become entrepreneurs with Θ_t^E.

3.3 Bank operating profit

Given the price of capital goods p_t, the wage w_t, the number of banks n_t, and its individual efficiency level, the expected *operating* profit of lending to the θ-type entrepreneurs is:

$$\pi^{B,O}(\theta; p_t, w_t, n_t, a_{jt}) = \frac{1-\xi}{n_t^2}\left(\eta\xi p_t \theta z_{t-1}\left(\frac{\left(1-\frac{1-\xi}{n_t}\right)a_{jt}}{(1+\tilde{\sigma})r^D w_t}\right)^{\xi}\right)^{\frac{1}{1-\xi}}. \tag{24}$$

Lending to more productive entrepreneurs is more profitable for banks, since they demand more funds. As p_t increases or w_t declines, the loan demand increases, which leads to higher profits for banks. Moreover, as either a_j or n_t increase, banks charge lower rate, which spurs entrepreneurs' borrowing and improves their profitability. As we discussed in Section 3.2, each bank holds a *portfolio* of lending contracts to all entrepreneurs in Θ_t^E. Integrating the bank profits given by Equation (24) over Θ_t^E, we obtain the bank's expected operating profit as:

$$\begin{aligned}\pi^{B,O}(\Theta_t^E; p_t, w_t, n_t, a_{jt}) &= M\int_{\theta\in\Theta_t^E}\pi^{B,O}(\theta; p_t, w_t, n_t, a_{jt})\,\mathbf{d}\,\mathbf{F}(\theta) \\ &= \frac{(1-\xi)M}{n_t^2}\left(\eta\xi p_t z_{t-1}\left(\frac{\left(1-\frac{1-\xi}{n_t}\right)a_{jt}}{(1+\tilde{\sigma})r^D w_t}\right)^{\xi}\right)^{\frac{1}{1-\xi}} G\left(\Theta_t^E\right),\end{aligned} \tag{25}$$

where M is the agent measure, $\pi^{B,O}(\theta; p_t, w_t, n_t, a_{jt})$ is defined in Equation (24), and $G\left(\Theta_t^E\right)$ is an aggregator of entrepreneurs' productivity given by:

$$G\left(\Theta_t^E\right) \equiv \int_{\theta\in\Theta_t^E}\theta^{\frac{1}{1-\xi}}\,\mathbf{d}\,\mathbf{F}(\theta). \tag{26}$$

Bank profitability is linked to the profitability of the capital good sector, as it improves when p_t rises relative to the cost of funds, r^D, or when w_t declines. Importantly, bank profitability depends on the number of banks n_t as well as the *range* of projects that are financed by the banking sector. The number of banks determines the market structure of the banking industry, which further affects banks' pricing power, consistent with a standard partial-equilibrium intuition under Cournot competition. As more highly productive agents choose to be entrepreneurs, the aggregator $G\left(\Theta_t^E\right)$ increases, which leads higher demand for loans, and thus makes banks more profitable.

3.4 Bank IT adoption

Before starting to lend, banks must invest in IT equipment and convert it into a bank-specific level of operating efficiency according to Equation (11). In doing so, banks can channel deposits to loans more efficiently. We denote the size of the IT equipment investment as q_{jt}. Now combing Equation (11) with (25), the total profit of bank j can be rewritten as

$$\pi^{B,O}(\Theta_t^E; p_t, w_t, n_t, q_{jt}) = \frac{(1-\xi)M}{n_t^2} \left(\eta \xi p_t z_{t-1} \left(\frac{1 - \frac{1-\xi}{n_t}}{(1+\tilde{\sigma}) r^D w_t} \right)^{\xi} \right)^{\frac{1}{1-\xi}} G\left(\Theta_t^E\right) \left(\kappa a_{t-1} q_{jt}^{\frac{1}{2}}\right)^{\frac{\xi}{1-\xi}} \tag{27}$$

Given the price of capital good p_t, the wage w_t, and the number of banks n_t, bank j chooses q_{jt} to maximize the profits as follows:

$$\pi^{B}(\Theta_t^E; p_t, w_t, n_t) = \max_{q_{jt}} \frac{(1-\xi)M}{n_t^2} \left(\eta \xi p_t z_{t-1} \left(\frac{1 - \frac{1-\xi}{n_t}}{(1+\tilde{\sigma}) r^D w_t} \right)^{\xi} \right)^{\frac{1}{1-\xi}} G\left(\Theta_t^E\right) \left(\kappa a_{t-1} q_{jt}^{\frac{1}{2}}\right)^{\frac{\xi}{1-\xi}} - p_t q_{jt}. \tag{28}$$

The first order condition for this problem is

$$\frac{\xi M}{2n_t^2} \left(\eta \xi p_t z_{t-1} \left(\frac{1 - \frac{1-\xi}{n_t}}{(1+\tilde{\sigma}) r^D w_t} \right)^{\xi} \right)^{\frac{1}{1-\xi}} G\left(\Theta_t^E\right) \left(\kappa a_{t-1} q_{jt}^{\frac{1}{2}}\right)^{\frac{\xi}{1-\xi}} = p_t q_{jt}, \tag{29}$$

which yields

$$q_{jt} = \frac{\xi}{2(1-\xi)} \frac{\pi^{B,O}(\Theta_t^E; p_t, w_t, n_t, q_{jt})}{p_t}. \tag{30}$$

This expression shows that IT adoption by banks increases as the banking sector is more profitable or capital goods become cheaper. The equilibrium total profit by bank j is

$$\pi^{B}(\Theta_t^E; p_t, w_t, n_t) = \frac{(2-3\xi)\, M}{2n_t^2} \eta \left(\frac{\xi p_t z_{t-1}}{\left(r_t^L w_t\right)^{\xi}} \right)^{\frac{1}{1-\xi}} G\left(\Theta_t^E\right). \tag{31}$$

The total profit responds to the change in the number of banks, the price of capital goods, the aggregate bank efficiency, the deposit rate, and the share of agents choosing to be entrepreneurs.

3.5 Equilibrium in the labor market

Agents who choose to be workers provide their labor, which is $M\left(1 - \mathbf{F}\left(\Theta_t^E\right)\right)$ units in total. Entrepreneurs hire workers in production. The market cleaning condition is

$$M\left(1 - \mathbf{F}(\Theta_t^E)\right) = M \int_{\theta \in \Theta_t^E} l(\theta)\, \mathbf{d}\, \mathbf{F}\left(\theta\right), \tag{32}$$

where the left hand side is the total labor supply, and the right hand size is the total labor demand. Combining labor demand from Equation (8) with (32), we have

$$w_t = \left(\frac{G\left(\Theta_t^E\right)}{1 - \mathbf{F}\left(\Theta_t^E\right)}\right)^{1-\xi} \frac{\xi p_t z_{t-1}}{r_t^L}. \tag{33}$$

As the labor supply increases, i.e. Θ_t^E shrinks, the wage decreases. As the price of capital goods increases, or as the interest rate charged by banks decreases, the labor demand increases and the wage rises.

3.6 Equilibrium in the capital good market

Entrepreneurs produce capital goods purchased by final good producers to use in production of final goods and by banks to improve their operating efficiency. Without aggregate uncertainty, p_t is determined by the market clearing condition:

$$\eta M \int_{\theta \in \Theta_t^E} \theta z_{t-1} (l_t(\theta))^{\xi}\, \mathbf{d}\, \mathbf{F}\left(\theta\right) = K_t + Q_t, \tag{34}$$

where K_t is defined in Equation (4), and Q_t is the total amount of capital goods used by banks, defined as

$$Q_t = n_t q_{jt} \tag{35}$$

Note that the left-hand side of Equation (34) is the aggregate supply of capital goods, and the right-hand side is the aggregate demand by final good producers and banks, and . Substituting Equations (4) and (30) into Equation (34), the price of the capital goods is given by:

$$p_t = \alpha^{\frac{1}{\alpha}} \left(\frac{M}{\xi}\left(1 - \frac{\xi^2}{2n_t}\right)\left(1 - \mathbf{F}\left(\Theta_t^E\right)\right) \eta r_t^L w_t\right)^{-\frac{1-\alpha}{\alpha}}. \tag{36}$$

Then combining Equations (33) and (36) we obtain

$$p_t = \alpha \left(\eta M \left(1 - \frac{\xi^2}{2n_t} \right) \left(1 - \mathbf{F}\left(\Theta_t^E\right) \right)^{\xi} \left(G\left(\Theta_t^E\right) \right)^{1-\xi} z_{t-1} \right)^{-(1-\alpha)}, \tag{37}$$

and

$$w_t = \frac{\alpha \xi}{\eta^{1-\alpha} r_t^L} \frac{\left(G\left(\Theta_t^E\right) \right)^{\alpha(1-\xi)} z_{t-1}^{\alpha}}{\left(M \left(1 - \frac{\xi^2}{2n_t} \right) \right)^{1-\alpha} \left(1 - \mathbf{F}\left(\Theta_t^E\right) \right)^{1-\alpha\xi}}. \tag{38}$$

From Equation (37) we can see that a higher level of aggregate labor productivity, $z_t - 1$, increases the production of capital goods and reduces its price, p_t. A higher number of banks leads to more competition that drives down their profits and reduce the banks' demand for capital goods, which price declines. The price of capital goods also depends on Θ_t^E. As more agents choose to become workers, i.e. Θ_t^E shrinks, the labor supply increases, reducing the wage and the cost of capital goods, which drives up the capital production. More production implies a lower price. In contrast, fewer entrepreneurs means fewer production of capital goods, which rises its price. As a result the effect is ambiguous.

From Equation (38) we can see that a higher lending rate, all else equal, lowers the wage, as entrepreneurs respond to higher borrowing costs by lowering demand and production. A rise in the number of banks reduces the banks' operating profits, which further reduces their demand for capital goods. It implies a lower wage as labor demand declines. As Θ_t^E shrinks, more agents choose to be a worker, and thus the wage drops as the labor supply increases.

Replacing the capital price and wage in Equation (8) respectively with Equation (37) and (38), we have the labor demand is given by

$$l(\theta) = \frac{1 - \mathbf{F}\left(\Theta_t^E\right)}{G\left(\Theta_t^E\right)} \theta^{\frac{1}{1-\xi}} \tag{39}$$

3.7 Banks free entry condition

Bank entry is pinned down by a free-entry condition as in Melitz (2003). Bankers keep entering as long as the total profits exceed their reserved value, i.e. $w_t \Gamma$. As more banks enter, the banking industry becomes more competitive, and the pricing power and total profits of banks declines. When the banks' total profit is equal to the reservation value, banks stop entering.

The banks free entry condition is

$$\frac{(2-3\xi)\,M}{2n_t^2}\eta\left(\frac{\xi p_t z_{t-1}}{\left(r_t^L w_t\right)^{\xi}}\right)^{\frac{1}{1-\xi}} G\left(\Theta_t^E\right)=w_t\Gamma \tag{40}$$

where the left hand side is j bank's total profit as defined in Equation (31). Substituting the price of capital goods and the wage respectively from Equation (37) and (38) into (40), we have

$$n_t=\sqrt{\frac{2-3\xi}{2}\frac{M}{\Gamma}\left(1-\mathbf{F}\left(\Theta_t^E\right)\right)\eta r_t^L}. \tag{41}$$

Thus, more bank enter and the industry becomes more competitive as the lending rate increases, or the share of agents that chose to be entrepreneurs, Θ_t^E, declines. Bank entry responds also to changes in the potential bankers' reservation values, and more entry occurs if either the wage or the reservation value of the potential entrants decline, as indicated in Equation (40). This provides another channel through which the real economy affects the equilibrium in the banking industry.

3.8 Occupation choice threshold

To provide intuition for the the trade-off that agents face in their occupation choice, let's consider two extreme cases. Suppose first that all agents choose to be workers. In this case, as Equation (37) shows, the price of capital goods is infinite, and the marginal benefit of being an entrepreneurs is infinitely large, which would lead some agents to become entrepreneurs. Analogously, when all agents choose to be entrepreneurs, Equation (38) implies that the wage goes to infinite, and being a worker earns more than what any entrepreneur can make. Therefore, it must be the case that some agents choose to become workers, while other to be entrepreneurs. Moreover, because all agents take as given the price of capital, the wage, and the number of banks, Equation (10) implies that the profits of entrepreneurs increase as their managerial ability increases. On the other hand, all workers earn the same wage. Thus we conjecture that there must exist a threshold level of managerial ability at which the marginal agent is indifferent between being an entrepreneur and a worker, such that any agent with a higher ability chooses to be an entrepreneur, and agents with a lower ability to be a workers.

Denote this threshold as $\hat{\theta}_t$. The subset of agents who choose to be entrepreneurs and workers are $\Theta_t^E = \left\{\theta \in \Theta : \theta > \hat{\theta}_t\right\}$ and $\mathbf{F}\left(\Theta_t^E\right) = 1 - \mathbf{F}\left(\hat{\theta}_t\right)$, respectively. Evaluated at the threshold level $\hat{\theta}_t$, Equations (26), (37), and (38), respectively, become:

$$
\begin{aligned}
G\left(\Theta_t^E\right) &= G\left(\hat{\theta}_t\right) = \int_{\hat{\theta}_t}^{\bar{\theta}} \theta^{\frac{1}{1-\xi}} \ \mathbf{d}\,\mathbf{F}(\theta), && (42)\\
p_t\left(\hat{\theta}_t\right) &= \alpha\left(\eta M\left(1-\frac{\xi^2}{2n_t}\right)\left(\mathbf{F}\left(\hat{\theta}_t\right)\right)^{\xi}\left(G\left(\hat{\theta}_t\right)\right)^{1-\xi} z_{t-1}\right)^{-(1-\alpha)}, && (43)\\
w_t\left(\hat{\theta}_t\right) &= \frac{\alpha\xi}{\eta^{1-\alpha} r_t^L\left(\hat{\theta}_t\right)} \frac{\left(G\left(\hat{\theta}_t\right)\right)^{\alpha(1-\xi)} z_{t-1}^{\alpha}}{\left(M\left(1-\frac{\xi^2}{2n_t}\right)\right)^{1-\alpha}\left(\mathbf{F}\left(\hat{\theta}_t\right)\right)^{1-\alpha\xi}}. && (44)
\end{aligned}
$$

Substituting the price of capital goods and the wage respectively from Equations (43) and (44) into (30), we have

$$
q_{jt}\left(\hat{\theta}_t\right) = \frac{\xi^2\eta}{2n_t^2} M z_{t-1}\left(\mathbf{F}\left(\hat{\theta}_t\right)\right)^{\xi}\left(G\left(\hat{\theta}_t\right)\right)^{1-\xi}. \tag{45}
$$

Next, replacing the a_j in Equation (21) using Equation (11) and (45), we obtain

$$
\eta r_t^L\left(\hat{\theta}_t\right) = \frac{(1+\tilde{\sigma})\, r^D}{\left(1-\frac{1-\xi}{n_t}\right)\kappa a_{t-1}\left(\frac{\xi^2\eta}{2n_t^2} M z_{t-1}\left(\mathbf{F}\left(\hat{\theta}_t\right)\right)^{\xi}\left(G\left(\hat{\theta}_t\right)\right)^{1-\xi}\right)^{\frac{1}{2}}}. \tag{46}
$$

Substituting the interest rate from Equation (41) with (46), we have

$$
n_t\left(\hat{\theta}_t\right) = 1 - \xi + \frac{2-3\xi}{2}\frac{M}{\Gamma}\frac{(1+\tilde{\sigma})\, r^D\left(\mathbf{F}\left(\hat{\theta}_t\right)\right)^{1-\frac{\xi}{2}}}{\xi\kappa a_{t-1}\left(\frac{\eta}{2} M z_{t-1}\left(G\left(\hat{\theta}_t\right)\right)^{1-\xi}\right)^{\frac{1}{2}}}. \tag{47}
$$

Replacing the price of capital goods from Equation (36) into (10), the profit of the type–θ entrepreneur becomes

$$
\pi^E\left(\theta,\hat{\theta}_t\right) = \left(\frac{1}{\xi}-1\right)\mathbf{F}\left(\hat{\theta}_t\right)\eta r_t^L\left(\hat{\theta}_t\right) w_t\left(\hat{\theta}_t\right)\frac{\theta^{\frac{1}{1-\xi}}}{G\left(\hat{\theta}_t\right)}. \tag{48}
$$

Proposition 2 states the existence of the equilibrium and show its uniqueness.

Proposition 2 (**Existence and Uniqueness of Equilibrium**). *Suppose that the labor share* *of capital goods production satisfies the following condition:*

$$\frac{2(1-\xi)^3}{(2-3\xi)\xi} < \frac{M}{\Gamma}\frac{G(\underline{\theta})}{\underline{\theta}^{\frac{1}{1-\xi}}}. \tag{49}$$

Then for any $\{a_{t-1}, z_{t-1}\}$*, there always exists a unique threshold* θ_t^**, where* $\pi^E(\theta_t^*, \theta_t^*) = w_t(\theta_t^*)$*, such that for* $\theta > \theta_t^*$*,* $\pi^E(\theta, \theta_t^*) > w_t(\theta_t^*)$*, whereas* $\pi^E(\theta, \theta_t^*) < w_t(\theta_t^*)$ *for* $\theta < \theta_t^*$*.*

Proof: see Appendix C.2.

To verify the existence of equilibrium, we show that the $\frac{\pi^E(\hat{\theta}_t, \hat{\theta}_t)}{w_t(\hat{\theta}_t)}$ is an increasing function of $\hat{\theta}_t$. Moreover, $\lim_{\hat{\theta}_t \to \infty} \frac{\pi^E(\hat{\theta}_t, \hat{\theta}_t)}{w_t(\hat{\theta}_t)} = \infty$ as $\lim_{\hat{\theta}_t \to \infty} G\left(\hat{\theta}_t\right) = 0$, whereas Assumption (49) guarantees that $\frac{\pi^E(\underline{\theta}, \underline{\theta})}{w_t(\underline{\theta})} < 1$. Therefore, there must exist a solution for $\frac{\pi^E(\hat{\theta}_t, \hat{\theta}_t)}{w_t(\hat{\theta}_t)} = 1$. Last, the uniqueness follows as the relative value is monotonic in $\hat{\theta}_t$. Now, we can rewrite $\pi^E(\theta_t^*, \theta_t^*) = w_t(\theta_t^*)$ as

$$\left(\frac{1}{\xi} - 1\right) \mathbf{F}(\theta_t^*)\, \eta r_t^L(\theta_t^*) \frac{(\theta_t^*)^{\frac{1}{1-\xi}}}{G(\theta_t^*)} = 1. \tag{50}$$

This implies that the threshold of occupation choice can be solved, using the following

$$\frac{2(1-\xi)}{(2-3\xi)\xi}\left(1-\xi+\frac{2-3\xi}{2}\frac{M}{\Gamma}\frac{(1+\tilde{\sigma})\, r^D (\mathbf{F}(\theta_t^*))^{1-\frac{\xi}{2}}}{\xi\kappa a_{t-1}\left(\frac{\eta}{2} M z_{t-1} (G(\theta_t^*))^{1-\xi}\right)^{\frac{1}{2}}}\right)^2 \frac{\Gamma}{M}\frac{(\theta_t^*)^{\frac{1}{1-\xi}}}{G(\theta_t^*)} = 1 \tag{51}$$

Equation (51) demonstrates how the existing economic and financial conditions, such as z_{t-1}, a_{t-1}, r^D, Γ and η, for instance, affect the threshold of occupation choice. The threshold, in turn, determines the number of banks, price of capital goods, the wage, and the interest rate. It is the critical link in the model connecting the real economy interacts with the financial system. As we will show in detail later, this threshold also affects the evolution of the aggregate technology z_t and the aggregate bank efficiency a_t, meaning that the growth of the real economy and the financial development also hinges on both the current economic and financial conditions.

3.9 Final goods market equilibrium

The market clearing conditions for the final goods market is

$$Y_t = C_t^F + C_t^E + C_t^W + C_t^B. \tag{52}$$

Since all agents are risk neutral, and assuming their discount rate is positive, they consume all their profits, so that the consumption of the final good producer is given by $C_t^F = \pi_t^F$, where π_t^F is the profit of final good producer, defined in Equation (5); the consumption of entrepreneurs is $C_t^E = M \int_{\theta_t^*}^{\bar{\theta}} \pi^E(\theta, \theta_t^*)\ \mathbf{d}\ \mathbf{F}(\theta)$; the consumption of workers is $C_t^W = w_t M \mathbf{F}(\theta_t^*)$; and the consumption of bankers is $C_t^B = n_t w_t \Gamma$.

4 Comparative statics

We now illustrate some of the model properties by looking at the response of the equilibrium number of banks, n_t, the occupation choice threshold, θ_t^*, and the model-based CBI to changes in the critical key endogenous state variables, (z_{t-1}, a_{t-1}), and exogenous drivers, (r^D, η).

4.1 The threshold of occupation choice and the number of banks

Equation (51) shows how the current economic and financial conditions determines the occupation choice threshold. The next proposition establishes the comparative statics with respect to changes in z_{t-1}, a_{t-1}, r^D, and η:[4]

Proposition 3 (**Occupation choice comparative statics**). *Suppose that θ_t^* is determined by Equation (51). Then, the following holds:*

$$(a) \qquad \frac{\mathbf{d}\ \theta_t^*}{\mathbf{d}\ z_{t-1}} > 0; \tag{53}$$

$$(b) \qquad \frac{\mathbf{d}\ \theta_t^*}{\mathbf{d}\ a_{t-1}} > 0; \tag{54}$$

$$(c) \qquad \frac{\mathbf{d}\ \theta_t^*}{\mathbf{d}\ r^D} < 0; \tag{55}$$

$$(d) \qquad \frac{\mathbf{d}\ \theta_t^*}{\mathbf{d}\ \eta} > 0. \tag{56}$$

Proof: see Appendix C.3.

Proposition 3 shows how changes in the level of the preexisting aggregate technology, the aggregate efficiency in the banking sector, the banks' funding cost level, and the probability of a project success affect the agents' occupation choice, which in turn generates complex dynamics.

[4]Comparative statics to all exogenous drivers can be easily established. We provide results for these particular ones as they are the structural parameters that we will shock to explain the time evolution of the CBI.

To provide the intuition, recall that the profit of the type–θ entrepreneur, from Equation (48), can be written as

$$\pi^E(\theta, \theta_t^*) = \left(\frac{1}{\xi} - 1\right) \underbrace{\eta r_t^L(\theta_t^*) w_t(\theta_t^*)}_{marginal\ cost\ of\ labor} \underbrace{\mathbf{F}(\theta_t^*) \frac{\theta^{\frac{1}{1-\xi}}}{G(\theta_t^*)}}_{labor\ demand}. \tag{57}$$

This expression is the product of three terms: some constant, the entrepreneur' marginal cost of labor, and the labor demand. While the worker's income equals the wage, the entrepreneur's marginal cost of labor depends on both the wage and the lending interest rate, as they need to finance wage from banks. Now, consider first an increase in z_{t-1}. This leads to increased production, as shown by Equation (10). More production generates a higher labor demand and pushes up the wage. However, more production also reduces the price of capital goods, which allow banks to increase IT investment to improve their efficiency and charge a lower lending interest rate. The lending rate declines, therefore, partially offsets the wage increase impact on the marginal cost. As a result, at the margin, the wage increases relatively more than the entrepreneur's profit. As a result more agents choose to become workers, i.e. θ_t^* increases.

Analogously, an increase in the aggregate banking efficiency allows banks to charge a lower rate, which lowers the borrowing cost for entrepreneurs. As a result they want to produce more and demand more labor, with an increasing wage. However, because the interest rate is now lower, the marginal cost of labor increases relatively less than the wage. Thus, more agents choose to be workers, i.e. θ_t^* increases.

Next, consider a rise in the deposit rate, r_D. A higher deposit rate translates into a higher lending rate, r_t^L, and borrowing costs for entrepreneurs, which discourages them to borrow and constraints production. Labor demand and the wage fall. Because the decline in the wage partially offsets the rise in the cost of funds for firms, the decline in the profit of the entrepreneur is smaller than the fall in the wage at the margin. As a result, being an entrepreneur becomes more attractive and θ_t^* decreases.

Last, consider a higher probability of success. A higher likelihood of succeeding increases the entrepreneurs' production, and thus demand for labor. The wage increases. Also, more production of capital goods implies a lower price, which leads banks to increase IT adoption, improving their efficiency and allowing them to charge a lower lending rate. This partially offset the impact of wage rise in the marginal cost of labor. As a result, again, the the wage increases relatively more than the the marginal cost of labor. Hence, being a worker becomes more attractive and θ_t^* increases.

Once we established the comparative statics of θ_t^*, it is easier to sign the response of the number of banks, n_t, and hence the market structure, to the same drivers. Replacing the term $\mathbf{F}\left(\theta_t^*\right)\eta r_t^L\left(\theta_t^*\right)$ in Equation (41) with (50), we obtain

$$n_t\left(\theta_t^*\right) = \sqrt{\frac{(2-3\xi)\xi}{2(1-\xi)}\frac{M}{\Gamma}\frac{G\left(\theta_t^*\right)}{\left(\theta_t^*\right)^{\frac{1}{1-\xi}}}}. \tag{58}$$

Equation (58) provides a direct channel through which the real economy affects the banking sector. As we saw earlier, bank entry depends on the threshold of occupation choice and the bankers' reservation value. A lower reservation value, which depends on the wage, leads to more bank entry. As θ_t^* increases, more agents choose to be workers and the wage rate declines to accommodate the higher labor supply. Also, the aggregate loan demand declines as fewer entrepreneurs want to produce. This reduces the banks' total profits and discourage bankers to enter. Moreover, the number of banks responds to the change in θ_t^* and Γ through two channels: the pricing power and the market share. As the number of banks increases, not only does the banks' pricing power decrease, but also the market share of each bank shrinks. The next collary formalizes these intuitions.

Collary 1 (Comparative statics of number of banks). *Suppose that θ_t^* and n_t are determined by Equation (51) and (58), respectively. Then we have the following comparative statics for the number of banks:*

$$(a) \qquad \frac{\mathbf{d}\, n_t}{\mathbf{d}\, z_{t-1}} < 0; \tag{59}$$

$$(b) \qquad \frac{\mathbf{d}\, n_t}{\mathbf{d}\, a_{t-1}} < 0; \tag{60}$$

$$(c) \qquad \frac{\mathbf{d}\, n_t}{\mathbf{d}\, r^D} > 0; \tag{61}$$

$$(d) \qquad \frac{\mathbf{d}\, n_t}{\mathbf{d}\, \eta} < 0. \tag{62}$$

Proof: see Appendix C.4.

Integrating Equation (9) over all entrepreneurs, the total loan supply issued by bank j is

$$B_{jt}\left(\theta_t^*\right) = \frac{M\int_{\theta_t^*}^{\bar{\theta}} b(\theta)\,\mathbf{d}\,\mathbf{F}\left(\theta\right)}{n_t\left(\theta_t^*\right)} = \frac{M}{n_t\left(\theta_t^*\right)}\left(\frac{\xi p_t\left(\theta_t^*\right) z_{t-1}}{r_t^L\left(\theta_t^*\right)\left(w_t\left(\theta_t^*\right)\right)^{\xi}}\right)^{\frac{1}{1-\xi}} G\left(\theta_t^*\right) = \frac{w_t\left(\theta_t^*\right) M\mathbf{F}\left(\theta^*\right)}{n_t\left(\theta_t^*\right)}, \tag{63}$$

where $w_t\left(\theta_t^*\right)$ and $n_t\left(\theta_t^*\right)$ are respectively defined in Equation (44) and (58), $b(\theta)$ is defined in Equation (9), and the third equality is due to Equation (33). Equation (63) reflects the fact that entrepreneurs borrow from banks to finance the wage bill. It also illustrates the connection between the total loans issued by individual bank j and the wage.

As we the bankers' reservation value is $w_t\Gamma$, Equation (63) permits us to compare the total bank profit and the reservation value. To see this, note that we can write any bank's total profit in Equation (31) as the product of four components: some constant, the Lerner index of banking industry, the risk-adjusted interest rate, and the total amount of loans by individual bank, as

$$\pi^B\left(\theta_t^*\right) = \frac{2-3\xi}{2(1-\xi)} \underbrace{\frac{1-\xi}{n_t\left(\theta_t^*\right)}}_{Lerner\ bindex} \underbrace{\eta r_t^L\left(\theta_t^*\right)}_{risk-adjusted\ rate} \underbrace{\frac{w_t\left(\theta_t^*\right) M\mathbf{F}\left(\theta^*\right)}{n_t\left(\theta_t^*\right)}}_{total\ loans}. \tag{64}$$

Note that the total amount of loans lent by an individual banks directly depends on the wage as the the total supply of labor is $M\mathbf{F}\left(\theta^*\right)$. First, a higher z_{t-1} or η directly increases the entrepreneur's output, which leads to a higher demand for credit and labor. As a result, the wage rises. However, more production in capital goods implies a lower price, which encourages banks to adopt IT, improving their efficiency, and lowering the lending interest rate. This implies that banks' total profit increase relatively less than the bankers' reservation value. Thus fewer bankers enter. Similarly, a rise in a_{t-1} improves banks' efficiency, which leads banks to lower the lending rate and further stimulating credit demand by entrepreneurs. Higher output of capital goods increases labor demand and the wage. Overall, the rise in the total profit of the bank is less than the increase in the reservation value. Thus, fewer bankers enter. Last, an increase in r^D rises the borrowing cost, which reduces the incentives to borrow and produce. The labor demand and the wage decline. However, a higher interest rate makes the decline in the total bank profit smaller than the contraction in the reservation value. Therefore, more bankers enter.

4.2 The CBI and its driving forces

Consistent with Philippon's (2015) definition of the CFI, we define and measure in the data the cost of banking intermediation (CBI) as the ratio of total bank net operating income over total intermediated assets, where the latter include cash, loans, securities and equity. The counterpart of the CBI in our model is the ratio of total profits to total bank loans, as banks in the model do not hold equity, cash, or securities and there is no labor employed by the banking sector. Define the *aggregate* profit of the banking sector as

$$\Pi^B\left(\theta_t^*\right) = n_t\left(\theta_t^*\right)\pi^B(\theta_t^*) = \frac{2-3\xi}{2n_t\left(\theta_t^*\right)}\eta r_t^L\left(\theta_t^*\right)w_t\left(\theta_t^*\right)M\mathbf{F}\left(\theta^*\right) \tag{65}$$

and the aggregate volume of loans intermediated by banks as

$$B_t\left(\theta_t^*\right) = n_t B_{jt}\left(\theta_t^*\right) = w_t\left(\theta_t^*\right)M\mathbf{F}\left(\theta^*\right) \tag{66}$$

where the second equality of Equation (65) is due to Equation (64), and the second equality in Equation (66) is due to Equation (63). The equilibrium CBI, denoted $\phi_t\left(\theta_t^*\right)$ in the model, is given by:

$$\phi_t\left(\theta_t^*\right) = \frac{\Pi^B\left(\theta_t^*\right)}{B_t\left(\theta_t^*\right)} = \frac{2-3\xi}{2(1-\xi)}\underbrace{\frac{1-\xi}{n_t\left(\theta_t^*\right)}}_{Lerner\ index}\underbrace{\eta r_t^L\left(\theta_t^*\right)}_{risk-adjusted\ rate} \tag{67}$$

where π^B and B_t are given by Equations (65) and (66), respectively. We can decompose the CBI into three components: some constant, the Lerner index in banking industry, and the risk-adjusted interest rate. Replacing the risk-adjust interest rate ηr_t^L from Equation (67) with (50), we obtain

$$\phi_t = \frac{1}{\mathbf{F}\left(\theta_t^*\right)}\sqrt{\frac{(2-3\xi)\xi}{2(1-\xi)}\frac{\Gamma}{M}\frac{G\left(\theta_t^*\right)}{\left(\theta_t^*\right)^{\frac{1}{1-\xi}}}} \tag{68}$$

The following proposition establishes how the CBI varies in response to changes in z_{t-1}, a_{t-1}, r^D, and η.

Proposition 4 (CBI comparative statics). *Given the CBI's definition in Equation (67), we* *have*

$$(a) \qquad \frac{\mathbf{d}\,\phi_t}{\mathbf{d}\,z_{t-1}} < 0; \tag{69}$$

$$(b) \qquad \frac{\mathbf{d}\,\phi_t}{\mathbf{d}\,a_{t-1}} < 0; \tag{70}$$

$$(c) \qquad \frac{\mathbf{d}\,\phi_t}{\mathbf{d}\,r^D} > 0; \tag{71}$$

$$(d) \qquad \frac{\mathbf{d}\,\phi_t}{\mathbf{d}\,\eta} < 0. \tag{72}$$

Proof: see Appendix C.5.

Equation (64) indicates that a rise in the number of banks directly reduces both the Lerner index of banking industry and the market share of any individual banks. First, consider an increase in z_{t-1} or η. This leads to more capital production, which reduces the capital price. This incentives banks to adopt more IT equipment and charge for a lower risk-adjusted interest rate. The lower rate directly reduce the CBI, and meanwhile has indirect effects on the CBI through the bank entry decision. In particular, a lower rate suggests that the bank total profit increases relatively less than the reservation value. Thus fewer banks enter, which is reflected by an increases in both their pricing power and the market share. However, only the part on the pricing power is counted in the CBI. Thus, the rise in the Lerner index only offset the part of the decline in the interest rate, which results a lower CBI.

Second, a rise in a_{t-1} makes banks monitoring more efficiently, and then charging for a lower interest rate, which further lowers the entrepreneurs' borrowing cost, and increases their demand for loans. As the rise in z_{t-1} or η, the reduce in interest rate has both direct and indirect effects on the CBI, and moreover the direct one dominates. Thus the CBI decreases as well.

In contrast, a rise in r^D leads to a higher lending rate that lowers credit demand and depresses output of capital goods, increasing their prices. More expensive IT equipment slows bank IT investment, curtails bank efficiency and increases the lending rate further. The higher rate directly jumps up the CBI, and also has indirect effects on the CBI by allowing more banks to enter, as the bank total profit decreases relatively less than the reservation value. Banks' pricing power and market share both drop. However, the CBI only counts the decline in the pricing power. Thus, the rise in the interest rate is relatively larger than the decline in the Lerner, which results a higher CBI.

To delve more into the behavior of the CBI as consequence of the joint evolution of the model endogenous and exogenous driving forces requires analyzing the long-run and short-run model dynamics, quantitatively, to which we now turn.

5 Balanced growth path

In this section, we discuss the model implied long-run dynamics. To begin with, we assume that the managerial ability of agents follows a geometric distribution, as in Axtell (2001) and Gabaix (2016), with the domain on $[\underline{\theta}, \infty)$ and CDF $\mathbf{F}(\theta) = 1 - \left(\frac{\underline{\theta}}{\theta}\right)^{\psi}$ and we will calibrate this CDF to match the distribution of the U.S. firm size, which roughly follows the 'Zipf's law'.

Next, we need to specify the rule of motion for for z_t and a_t. Following Melitz (2003), we posit that z_t evolves according to

$$z_t = \frac{z_{t-1}}{1 - \mathbf{F}(\theta_t^*)} \int_{\theta_t^*}^{\infty} \eta\theta \, \mathbf{d}\, \mathbf{F}(\theta). \tag{73}$$

Thus, period t gross rate of growth of z_t is given by the average expected managerial ability across all agents that choose to be entrepreneurs and evolves endogenously, conditional on the threshold value for the occupation choice and the preexisting level of aggregate technology that affects current decisions. Solving this integral, we obtain:

$$g_t^Z = \frac{z_t}{z_{t-1}} = \frac{\eta\psi}{\psi - 1}\theta_t^*. \tag{74}$$

Equation (51) suggests that the threshold of occupation choice, θ_t^*, depends on both the preexisting level of aggregate technology, z_{t-1}, and the preexisting level of aggregate banking efficiency, a_{t-1}. Thus the evolution of θ_t^* is jointly determined by z_{t-1} and a_{t-1}. Importantly, aggregate technology growth, g_t^Z, since it depends on θ_t^* that solves Equation (51), not only depends on the structural parameters and other endogenous variables in the model, but also on a_{t-1}. Thus, there is a channel in the model through which financial sector development, as captured by a_{t-1}, can affect the economy's long-run growth and we average across all banks in the market.

To specify the evolution of aggregate banking efficiency in a manner consistent with Equation (73), we take three steps. First, we rewrite the individual efficiency a_{jt} from bank j, as defined in Equation (11), substituting q_{jt} from Equation (45) and n_t from Equation (58), and use

$\mathbf{F}(\theta) = 1 - \left(\frac{\theta}{\underline{\theta}}\right)^{\psi}$ to obtain

$$
\begin{aligned}
a_{jt} &= \kappa a_{t-1} \left(\frac{(1-\xi)\xi\eta}{2-3\xi} z_{t-1} \Gamma \left(\frac{\mathbf{F}(\theta_t^*)}{G(\theta_t^*)} \right)^{\xi} (\theta_t^*)^{\frac{1}{1-\xi}} \right)^{\frac{1}{2}} \\
&= \kappa a_{t-1} \left(\frac{(1-\xi)^{1-\xi}\xi\eta}{2-3\xi} \left(1 - \xi - \frac{1}{\psi}\right)^{\xi} \Gamma \left(\left(\frac{\theta_t^*}{\underline{\theta}} \right)^{\psi} - 1 \right)^{\xi} \theta_t^* \right)^{\frac{1}{2}} z_{t-1}^{\frac{1}{2}}, \qquad (75)
\end{aligned}
$$

where $G(\theta_t^*)$ is defined in Equation (42). Second, we average Equation (75) across all banks. Third, to induce well-behaved long-run dynamics and obtain a BGP, we make two adjustments. First, we scale this average relative to the risk-adjusted non-financial sector productivity level, $\eta\sqrt{z_{t-1}}$, in order to remove the direct effect of these two factors from the evolution of the level of aggregate banking efficiency. Second, and finally, we weight the average relative efficiency with the factor, $\tau \left(\frac{1-\mathbf{F}(\theta_t^*)}{\mathbf{F}(\theta_t^*)}\right)^{\delta}$, where $\frac{1-\mathbf{F}(\theta_t^*)}{\mathbf{F}(\theta_t^*)}$ is the entrepreneur share relative to the worker share, and $\tau > 0$ and $\delta > 0$ are two scaling parameters.

While this last adjustment term is needed to obtain a well-behaved BGP dynamics, it can be be interpreted as capturing the role of organization capital– e.g., Atkeson and Kehoe (2005) and McGrattan and Prescott (2005). To improve their efficiency, not only do banks adopt IT equipment, but also accumulate soft skills, like experience in running their changing business models, information processing, marketing, branding, etc. This adjustment term implies that these skills' stock increases as the threshold declines and banks lend to a broader pool of entrepreneurs. The parameters τ and δ control the level and the strength of this effect.

The resulting expression for aggregate banking efficiency is:

$$
a_t = \tau \left(\frac{1-\mathbf{F}(\theta_t^*)}{\mathbf{F}(\theta_t^*)} \right)^{\delta} \frac{1}{n(\theta_t^*)} \sum_{j=1}^{n(\theta_t^*)} \frac{a_{jt}}{\eta\sqrt{z_{t-1}}}, \qquad (76)
$$

which implies that the gross rate of growth of aggregate banking sector efficiency evolves according to:

$$
g_t^A = \frac{a_t}{a_{t-1}} = \tau\kappa \left(\frac{(1-\xi)^{1-\xi}\xi}{(2-3\xi)\eta} \left(1 - \xi - \frac{1}{\psi}\right)^{\xi} \Gamma\theta_t^* \right)^{\frac{1}{2}} \left(\left(\frac{\theta_t^*}{\underline{\theta}} \right)^{\psi} - 1 \right)^{\frac{\xi}{2}-\delta}. \qquad (77)
$$

Aggregate banking efficiency growth, therefore, depends on the threshold for the occupation choice, θ_t^*, as well as several structural parameters of the economy. As a result, developments on the real side of the economy can affect financial development. So, we can now characterize the balanced growth path of our economy with the following two propositions.

Proposition 5 (**Balanced growth path existence and uniqueness**). *Suppose that $\tilde{\theta}^*$ solves*

$$\tau\kappa\left(\frac{(1-\xi)^{1-\xi}\xi\psi}{(2-3\xi)(\psi-1)}\left(1-\xi-\frac{1}{\psi}\right)^{\xi}\Gamma\right)^{\frac{1}{2}}\tilde{\theta}^*\left(\left(\frac{\tilde{\theta}^*}{\underline{\theta}}\right)^{\psi}-1\right)^{\frac{\xi}{2}-\delta}=1, \tag{78}$$

and that

$$\delta>\frac{\xi}{2}+\frac{1}{\psi}. \tag{79}$$

Then, for any $\{\tilde{z}_0,\tilde{a}_0\}$ satisfying

$$\tilde{a}_0^2\tilde{z}_0=\frac{2(1-\xi)}{(2-3\xi)\xi}\left(1-\xi+\frac{2-3\xi}{2}\frac{M}{\Gamma}\frac{(1+\tilde{\sigma})\,r^D\left(\mathbf{F}\left(\tilde{\theta}^*\right)\right)^{1-\frac{\xi}{2}}}{\xi\kappa\left(\frac{\eta}{2}M\left(G\left(\tilde{\theta}^*\right)\right)^{1-\xi}\right)^{\frac{1}{2}}}\right)^2\frac{\Gamma}{M}\frac{\left(\tilde{\theta}^*\right)^{\frac{1}{1-\xi}}}{G\left(\tilde{\theta}^*\right)}, \tag{80}$$

there exists a unique balanced growth path such that

1. *the threshold value for the occupation choice is constant at $\tilde{\theta}^*$;*
2. *the aggregate level of non-financial productivity grows at the gross rate $g^Z=\frac{\eta\psi}{\psi-1}\tilde{\theta}^*$;*
3. *the aggregate level of banking efficiency declines at the gross rate of $g^A=(g^Z)^{-\frac{1}{2}}$, with $g^A=\tau\kappa\left(\frac{(1-\xi)^{1-\xi}\xi}{(2-3\xi)\eta}\left(1-\xi-\frac{1}{\psi}\right)^{\xi}\Gamma\tilde{\theta}^*\right)^{\frac{1}{2}}\left(\left(\frac{\tilde{\theta}^*}{\underline{\theta}}\right)^{\psi}-1\right)^{\frac{\xi}{2}-\delta}$;*
4. *the number of banks is constant at $n\left(\tilde{\theta}^*\right)=\sqrt{\frac{(2-3\xi)\xi}{2(1-\xi)}\frac{M}{\Gamma}\frac{G\left(\tilde{\theta}^*\right)}{\left(\tilde{\theta}^*\right)^{\frac{1}{1-\xi}}}}$;*
5. *the CBI is constant at $\phi\left(\tilde{\theta}^*\right)=\frac{1}{\mathbf{F}\left(\tilde{\theta}^*\right)}\sqrt{\frac{(2-3\xi)\xi}{2(1-\xi)}\frac{\Gamma}{M}\frac{G\left(\tilde{\theta}^*\right)}{\left(\tilde{\theta}^*\right)^{\frac{1}{1-\xi}}}}$;*
6. *the loan rate is constant at $r^L\left(\tilde{\theta}^*\right)=\frac{\xi}{(1-\xi)\eta}\frac{G\left(\tilde{\theta}^*\right)}{\mathbf{F}\left(\tilde{\theta}^*\right)\left(\tilde{\theta}^*\right)^{\frac{1}{1-\xi}}}$;*
7. *the individual bank efficiency is constant at $a_j\left(\tilde{\theta}^*\right)=\kappa\tilde{a}_0\left(\frac{(1-\xi)\eta}{2-3\xi}\tilde{z}_0\left(\tilde{\theta}^*\right)^{\frac{1}{1-\xi}}\left(\frac{\mathbf{F}\left(\tilde{\theta}^*\right)}{G\left(\tilde{\theta}^*\right)}\right)^{\xi}\right)$*

8. *the aggregate output, Y_t, the consumption by all agents, $\{C_t^F, C_t^E, C_t^W, C_t^B\}$, the aggregate volume of credit, B_t, the wage rate, w_t, and the profit of all agents, $\left\{\pi_t^f, \pi_t^E, \pi_t^B\right\}$, all grow at the gross rate $(g^Z)^\alpha$;*

9. *capital goods respectively used by the final good producers and banks, K_t and Q_t, grow at the gross rate of g^Z, while the the price of capital goods, p_t, declines at the gross rate $(g^Z)^{-(1-\alpha)}$.*

Proof: see Appendix C.6.

Proposition 5 states that, if the economy start at some initial point $\{\tilde{z}_0, \tilde{a}_0\}$, where $\tilde{\theta}^*$ solves Equation (51) at $\{\tilde{z}_0, \tilde{a}_0\}$, then there is a unique balanced growth path of the economy along which the threshold for the occupation choice, the number of banks, the CBI, and the interest rate remain constant, while the real economy expands, with different variables growing at different rates. The constant threshold value for the occupation choice, $\tilde{\theta}^*$, and hence the balance growth path of the economy, depends on the labor share in the production of capital goods, ξ, the parameters of the managerial ability distribution, ψ and $\underline{\theta}$, the bankers' reservation value parameter, Γ, and three scalg parameters τ, κ, and δ.

Note here that, to obtain a unique BGP, aggregate banking sector efficiency must decline at a rate that partially offsets the higher level of technological progress. In other words, as the economy grows due to technological advances in the non-financial sector, aggregate banking sector efficiency must play a declining role to guarantee the existence of a BGP. This BGP condition requirement is imposed by Assumption (79).[5] This assumption firstly ensures that Equation (78) has a unique solution. Second, it makes sure that the growth rate of $a_{t-1}\sqrt{z_{t-1}}$ is decreasing in θ_t^*. The intuition is that the organization capital component of aggregate banking efficiency has to play a progressively smaller role than IT capital accumulation in the efficiency of individual banks as the economy evolves.

The balanced growth path of the economy not only exists and is unique but, under further parametric restrictions on the rate of decline of aggregate banking sector efficiency, is also stable with the economy always reverting to it after an exogenous shock, as Proposition 6 illustrates.

[5]The assumption could be relaxed at the cost of loosing uniqueness and, as the next proposition shows, the stability of the BGP. It is not needed to prove existence of the BGP. Uniqueness and stability of the BGP are desirable properties to analyze the model's ability to interpret the data. This further means that the dynamics in the data that we want to explain with the model could depend by initial conditions.

Proposition 6 (**Balanced growth path stability**). *Under Assumption (79), for any given* $\{z_0, a_0\}$, θ_t^* *will converge to* $\tilde{\theta}^*$, *where* $\tilde{\theta}^*$ *is defined in Equation (78).*

Proof: see Appendix C.7.

Proposition 6 implies that, given any initial state, $\{z_0, a_0\}$, the threshold for the occupation choice will always revert to the same value determined by Equation (78). Thus, if $\theta_t^* < \tilde{\theta}^*$, $a_{t-1}\sqrt{z_{t-1}}$ grows at a higher rate than in the BGP in Proposition 5 and θ_t^* will increase. Thus, the threshold keeps rising until it reaches $\tilde{\theta}^*$. On the other hand, if $\theta_t^* > \tilde{\theta}^*$, $a_{t-1}\sqrt{z_{t-1}}$ contracts and the next period threshold falls, continuing to decline until it reaches $\tilde{\theta}^*$.

6 The long-run value of the CBI and its determinants

In this section we calibrate the balanced growth path of our model economy to match the long-run value of the CFI and basic stylized facts the U.S. economy growth. We then use the calibrated model to investigate the CBI drivers and its implications for long-run economic growth.

6.1 Calibration

The parameters that we need to calibrate are $\left\{\alpha,\ \eta,\ r^D,\ \xi,\ \sigma,\ \Gamma,\ \psi,\ \underline{\theta},\ \tau,\ \kappa,\ \delta\right\}$. The observable variables that we target are the following: the long-run value of the CFI, constant at 1.88% as in Philippon (2015); a annual average labor productivity growth from 1960-2015 of 2.1% from the BLS; a share of agents who chose to be workers, set to 70%; a long-run deposit-to-loans ratio in the U.S. banking system, arbitrarily set at 0.99; the coefficient of the logarithm of the firm size regressed on the logarithm of the corresponding frequency, set to -2.059 following Axtell (2001) and Gabaix (2016); and a lending-deposit interest rate spread set at 8%, corresponding to the average value in the Call Report data from 1985 to 2015.

Table 1 summarizes the resulting calibration. The capital share in the production of final goods, α, has a standard value of 0.38. We set $\eta = 0.83$ to match the one-year U.S. start up survival rate, between 1980 and 2014, using the Business Dynamics Statistics collected by the U.S. Census Bureau. We then choose a real deposit rate of 1% following Drechsler, Savov and Schnabl (2017).

As documented in Axtell (2001) and Gabaix (2016), the slope of the logarithm of the U.S. firm size distribution regressed on the frequency is about -2.059. Using Equation (39), we can

Table 1: Calibration Summary

	Parameter	Value	Target
α	capital share in final goods sector	0.38	capital share
η	prob of realizing managerial ability	0.83	startup survival rate
r^D	deposit rate	1.01	Drechsler et al (2017)
ξ	labor share in capital goods sector	0.47	9% risk-adjusted rate
σ	scalar for prob of finding true state	0.24	8% interest rate spread
Γ	reservation value	$7.59E(-4)$	1.88% Long-run CFI
ψ	distribution curvature	2.86	firm size distribution
$\underline{\theta}$	lower bound for managerial ability	0.54	share of worker
κ	scale factor for agent efficiency	70.76	deposit–to–loans ratio
τ	scale factor for growth rate of bank efficiency	1.36	$g^a\sqrt{g^Z} = 1$ along with BGP
δ	elasticity of adjustment term	0.6	Proposition 6
M	agent measure	1.0	normalized to 1

express the labor hired by the type-θ entrepreneur as

$$\ln l(\theta) = \ln\left(1 - \frac{1}{(1-\xi)\psi}\right) + \ln\left(\left(\frac{\theta_t^*}{\underline{\theta}}\right)^{\psi} - 1\right) + \frac{1}{1-\xi}\ln\theta. \tag{81}$$

Moreover, given $\mathbf{F}(\theta)$ and θ_t^*, we can calculate the logarithm of the share of the type–θ entrepreneur as

$$\ln h(\theta) = \ln\left(\frac{f(\theta)}{1-\mathbf{F}(\theta_t^*)}\right) = \psi\ln(\theta_t^*) + \ln\psi - (\psi+1)\ln\theta, \tag{82}$$

where $f(\theta)$ is the p.d.f. of the managerial ability distribution. Using results from Axtell (2001) and Gabaix (2016), we have

$$1 + \psi = \frac{2.059}{1-\xi}. \tag{83}$$

Next, we can rewrite Equation (50) as

$$\left(\frac{1}{\xi} - 1\right)\eta r_t^L = \frac{G(\theta_t^*)}{\mathbf{F}(\theta_t^*)(\theta_t^*)^{\frac{1}{1-\xi}}} = \frac{\psi}{\psi - \frac{1}{1-\xi}}\frac{1}{\left(\frac{\theta_t^*}{\underline{\theta}}\right)^{\psi} - 1}. \tag{84}$$

Here, we set the risk-adjusted interest rate, ηr_t^L at 9%, and the fraction of agents who are workers at 0.7, which implies $\frac{1}{\left(\frac{\theta_t^*}{\underline{\theta}}\right)^{\psi}-1} = \frac{1-0.7}{0.7} = \frac{3}{7}$. Thus combining Equation (83) and (84), we have $\psi = 2.86$ and $\xi = 0.47$. Note that $\xi < \frac{2}{3}$, satisfying the restriction we impose on ξ in Equation (6).

Equation (32) shows that the total labor supply is $M\mathbf{F}(\theta_t^*)$, whereas the aggregate capital good output is $\eta M \int_{\theta_t^*}^{\infty} \theta z_{t-1}(l_t(\theta))^{\xi} \ \mathbf{d} \ \mathbf{F}(\theta)$, according to Equation (34). Therefore, the aggregate labor productivity implied by our model is given by

$$z_t^L = \frac{\eta \int_{\theta_t^*}^{\infty} p_t \theta z_{t-1}(l_t(\theta))^{\xi} \ \mathbf{d} \ \mathbf{F}(\theta)}{\mathbf{F}(\theta_t^*)} = \alpha\xi\eta^{\alpha} \frac{(G(\theta_t^*))^{\alpha(1-\xi)} z_{t-1}^{\alpha}}{\left(M\left(1-\frac{\xi^2}{2n_t(\theta_t^*)}\right)\right)^{1-\alpha} (\mathbf{F}(\theta_t^*))^{1-\alpha\xi}}, \tag{85}$$

where the second equality is due to Equations (43) and (44). Equation (85) implies that, along the BGP, the gross rate of growth of labor productivity is constant and linked to z_t. We set $g^{LP} = \frac{z_t^L}{z_{t-1}^L} = (g^Z)^{\alpha} = 1.021$, which implies $g^Z = (g^{LP})^{\frac{1}{\alpha}} = 1.055$.[6] Given that there is no population growth in the model, this implies that real GDP growth along the BGP per capita is also 2.1%, which is only slightly higher than what we observe in data provided by the U.S. Bureau of Economic Analysis, 1.9% over the sample period. Equation (74) implies that the threshold for the occupation choice along the BGP is 0.83. Combined with the worker share of agents that we assumed to be 70%, we obtain $\underline{\theta} = 0.54$.

Finally, after normalizing the agent measure to 1, the CBI target value of 1.88% in the BGP combined with Equation (68) pins down $\Gamma = 7.59 \times 10^{-4}$. A bank-loan-to-deposit ratio of 0.99 determines $\kappa = 70.45$. Using Equation (21), we set $\sigma = 0.24$ to match the assumed 8% spread between lending and deposit rate. Now, Proposition 6 requires that $\delta > \frac{\xi}{2} + \frac{1}{\psi} = 0.58$. Thus we set $\delta = 0.60$. We finally note that Proposition 5 implies that $g_t^A\left(\tilde{\theta}^*\right)\left(g_t^Z\left(\tilde{\theta}^*\right)\right)^{\frac{1}{2}} = 1$, where $g_t^Z\left(\tilde{\theta}^*\right)$ and $g_t^A\left(\tilde{\theta}^*\right)$ are respectively defined in Equation (74) and (77). This yields $\tau = 1.36$.

6.2 CBI, imperfect competition and costly state verification

We are now ready to evaluate the structural determinants of the CBI. There are two main frictions in the model: imperfect competition and costly state verification. To evaluate their relative importance, we shut down the monitoring cost, by setting $\sigma = 0$ in Equation 12. Under this assumption, banks can always verify the true state, regardless the level of exerted efforts, and the component of the interest spread due to costly state verification vanishes. We then recalculate the implied CBI along the BGP. This counterfactual shows that the CBI is mainly

[6] Note that the aggregate technology, z_t, should not be interpreted as TFP in the model, since labor is the only input in production of capital goods. Growth of any factors other than labor, including capital, TFP, and utilization, etc., can contribute to z_t growth. As a result, z_t does not have a direct counterpart in the data. However, we can explicitly calculate the model implied labor productivity that we use to back out a model-implied growth rate for z_t.

driven by the imperfect competition, as the resulting value of the new CBI is 96.5% of our target value. The result reflects the fact that monitoring costs have a small contribution to the interest rate spread in our baseline calibration. Thus, one important result of the paper is that imperfect competition is the main driver of the CBI in our model.

Nonetheless, the long-run growth gain from eliminating the information asymmetry is sizable. Without costly state verification, the interest rate spread is 6.23 percentage points lower, declining from 8% to 1.77%. With a lower borrowing cost, entrepreneurs produce more and demand more labor and credit, which leads to a higher wage. However, the marginal cost of labor increases less than the wage increase as the lending rate declines. Hence more agents choose to become workers, the average managerial ability increases, leading to higher aggregate technology and labor productivity growth. In particular, the long-run growth rate of the aggregate labor productivity, implicitly defined by Equation (85), increases by 0.6 percentage points (almost 30 %), from 2.1% to 2.7%.

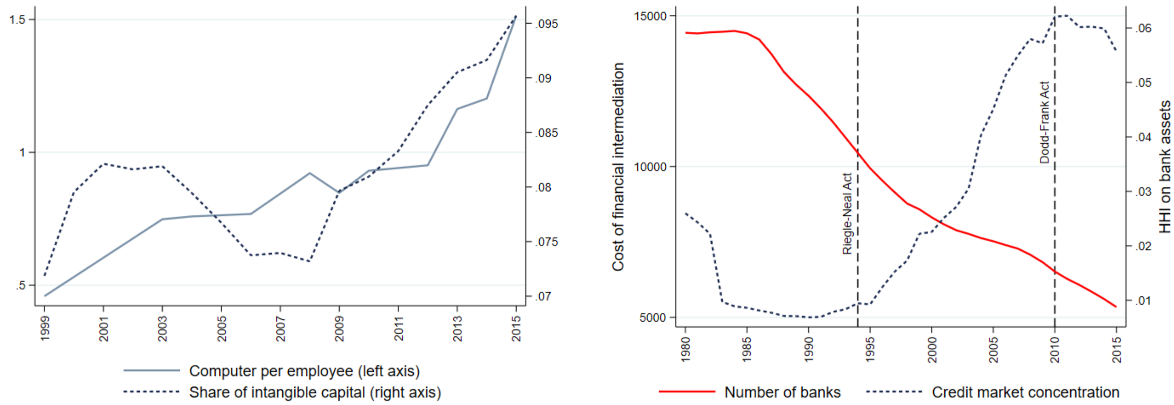
We interpret these results as suggesting that imperfect competition is the most important determinant of the CBI. In our set up without labor in banking, under perfect competition and no costly state verification the CBI vanishes. With labor in banking, it would be positive also under perfect competition. In the data, however, only the profit component of the value added in banking closely tracks the CFI, while the aggregate wage component of the banking value added is more or less constant over time (not reported).

7 Tracking the evolution of the CFI since mid-1960s

In this section, we explore whether our model can account for the inverted U-shape dynamics of the CFI between 1965 and 2015 in Figure 1. Before proceeding it is useful to discuss the evolution in the data of the variables that drive the CFI comparative statics in our model and the other variables that we will use to validate the mechanism proposed.

The main driver of the CBI in the model is aggregate non-financial technology that we match to the data through the model implied labor productivity. In the model, the state of the aggregate banking efficiency in the banking sector, a_{t-1}, also drives the CBI. As there is no commonly accepted empirical measure of aggregate bank efficiency, we do not consider exogenous changes in a_{t-1}. In all our simulations, a_{t-1} is left unrestricted and always evolves endogenously. A second possible CBI driving force, in the model, is the bank funding cost, r^D, which during

Figure 3: Bank IT investment and market structure



(a) Intangible capital and computer per employee

(b) Number of banks and market concentration

Notes: Panel A plots bank intangible assets as a share of total non-residential private assets and the number of computer per worker for banks over the period 1999-2015. Data are from the Bureau of Economic Analysis Fixed Assets Tables and from CiTBDs Aberdeen (Pierri and Timmer, 2020). Panel B plots the number of commercial banks and the Herfindahl-Hirschman Index computed on bank assets over the period 1980-2015. The vertical lines are drawn in correspondence of the two important changes in regulation: the Riegle-Neal and the Dodd-Frank Acts. Data are from FDIC and U.S. Call Reports.

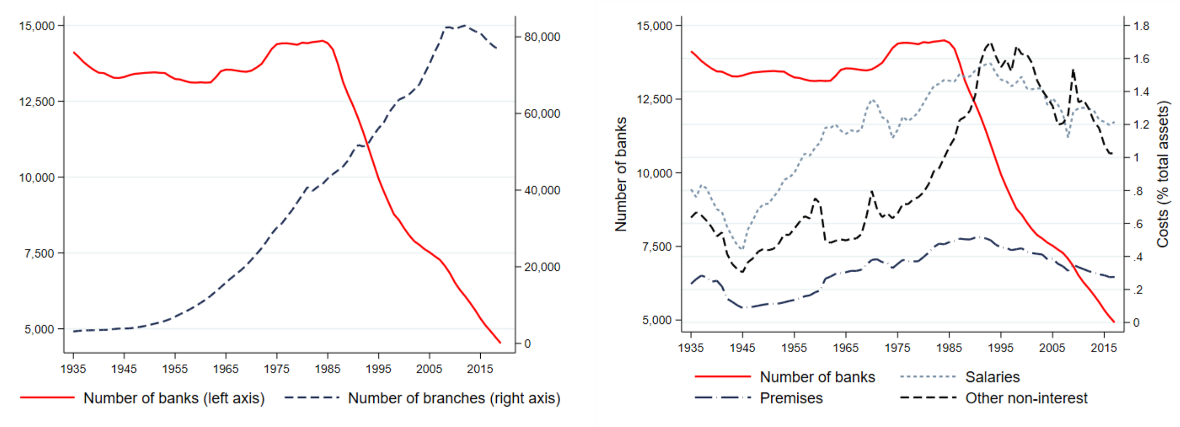
the period that we consider fluctuated sharply (Drechsler, Savov and Schnabl, 2020). Thus, we will explore the behavior of the CBI in response to changes in r^D and aggregate productivity growth that can be clearly identified in the data.

A similar issue arises in choosing the empirical counterparts of the variables that we use to validate the model mechanisms, in addition to the CBI. A variable that we can more confidently measure in the data is bank investment in IT equipment. We proxy the model concept of investment in IT equipment, Q_t, with with bank investment in intangible capital as a share of total non-residential private assets from the U.S. BEA, as in Philippon and Reshef (2013). Panel a of Figure 3 illustrates that this variable correlates closely with the number of computer per worker, an even closer proxy for IT-adoption in banking (Pierri and Timmer, 2020).

Assessing empirically the evolution of the banking sector market structure is particularly challenging. We relate the model's notion of a "bank" to the number of commercial banks in the data from the U.S. FDIC. Panel b of Figure 3 shows that a standard asset-based measure of bank concentration–the Herfindahl-Hirschman Index computed on bank assets (henceforth, HH Index)–correlates very closely with the number of commercial banks.

Even though we do not consider the timing of specific regulatory interventions potentially

Figure 4: Banks, branches, and cost structure



(a) Number of banks and branches

(b) Cost structure

Notes: The chart plots the number of commercial banks and branches (Panel a) and the share of rents, wages and other costs over bank assets (Panel b), from 1935 to 2017. Source: FDIC.

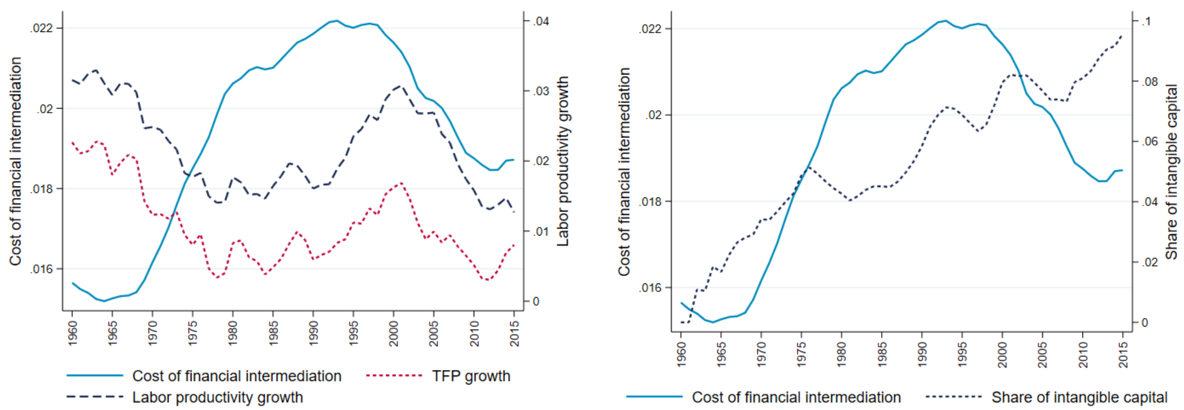
affecting bank entry, Figure 4 reports the total number of branches (Panel a) and other cost indicators (Panel b) for U.S. commercial banks that can capture the fixed costs of setting up a new bank. These cost indicators also correlates closely with the number of banks, although a change in the number of banks could of course be driving the aggregate dynamics.

7.1 Candidate drivers and outcome variables

Figure 5 plots the 10-year moving average of the CFI from 1965 to 2015 together with the level of U.S. TFP and labor productivity growth (Panel a), as well as the share of intangible capital in total non-residential private assets in banking (Panel b). The figure shows that the beginning of the productivity slowdown of the U.S. economy, in the mid-1960s, coincides with the first turning point in the CFI, while the subsequent partial recovery and acceleration roughly coincides with the flattening of the CBI in the late-1980s and its subsequent decline. This association weakens in the last part of the sample and depends also on the width of the moving average window, with a closer association when we use a 20-year window (not reported). The figure also suggests that intangible investment in banking, which we take as a proxy for banking industry IT adoption, tracks productivity closely but with some delay (Panel b).

Panel a of Figure 6 plots the 10-year average of the CFI together with a proxy for the real the deposit rate. Consistent with the evidence in Drechsler, Savov and Schnabl (2020), it shows

Figure 5: CFI, PRODUCTIVITY SLOWDOWN, AND INTANGIBLES



(a) CFI and TFP growth

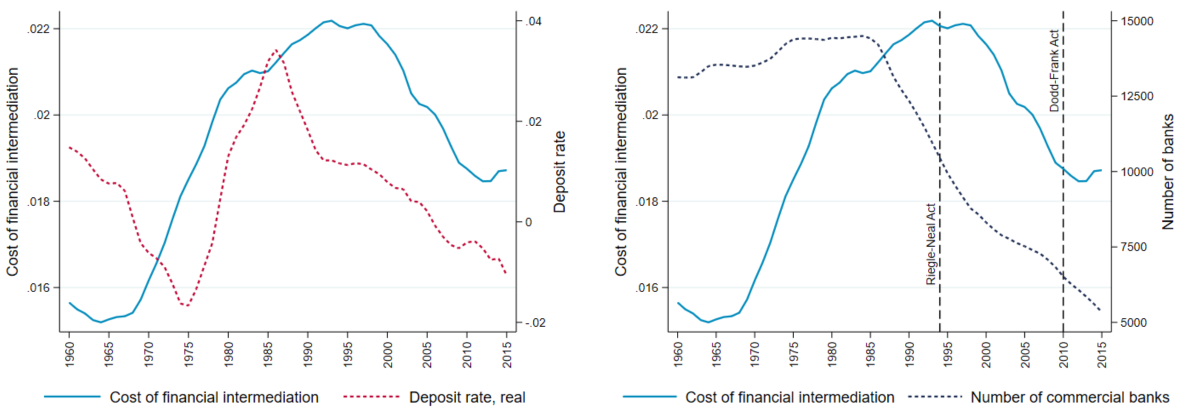
(b) CFI and intangible capital

NOTES: the charts plot the 10-year moving average of the CFI together with labor and total factor productivity growth (Panel a) and bank intangible assets as a share of total non-residential private assets (Panel b) over the period 1960-2015. Data are from Fernald (2012) and the Bureau of Economic Analysis Fixed Assets Tables.

that the bank funding costs declined through the end of the 1980s due increasing inflation and a binding cap on the nominal deposit rates imposed by Regulation Q. It increased very sharply for a period of about five years in the late-1970s and early-1980s due to the gradual removal of Regulation Q and the Volcker disinflation. It then declined steadily thereafter. While the turning points in this measure of bank funding costs are not as closely aligned with those in the CBI as in the case of productivity, this variable is a plausible contributor to the CBI low frequency dynamics according to Proposition 4.

Panel b of Figure 6 plots the number of commercial banks. It illustrates the well-known fact that the number of banks kept steady or increases slightly from 1955 to the mid-1980s, and then declines markedly thereafter with the deregulation process that started in the early 1970s culminating with the Riegle-Neal Interstate Banking and Branching Efficiency Act that lifted the remaining restrictions on branching within and across state borders. The declining trend in the number of banks is reflected in increased credit market concentration (Cetorelli, Hirtle, Morgan, Peristiani and Santos, 2007; McCord, Prescott and Sablik, 2015). Note in particular the coincidence of the turning points in the number of banks and the beginning of the plateau of the CBI in 1985. As Figure 3 (Panel b) shows, the decline in the number of banks is accompanied by an increase in the HH index of market concentration after the completion of interstate branch

Figure 6: CFI, funding costs, and number of banks



(a) CFI and the real deposit rate

(b) CFI and number of banks

Notes: The figure plots the 10-year moving average of the CFI together with the real deposit rate (Panel a) and the number of banks (Panel b) over the period 1960-2015. The real deposit rate is calculated deflating the nominal deposit rate computed from U.S. Call Reports with the U.C. CPI. Data are from U.S. Call Reports, FRED Economic Data and FDIC. In Panel b the vertical lines are drawn in correspondence of the two important changes in regulation: the Riegle-Neal and the Dodd-Frank Acts.

deregulation process in the mid-1990s, which halted after the Dodd-Frank Act in 2010.[7]

7.2 Two counterfactual scenarios and the CFI

To evaluate our model's ability to reproduce the inverted U-shape behavior in the U.S. CFI since the mid-1960s, we assume that the economy starts in 1965 from the balanced growth path calibrated above, initializing the system by arbitrarily setting $a_0 = 1$ and $z_0 = 1$. Thus, we do not take a stance on the sources of the persistent decline in the CFI from the 1920s to mid-1960s, nor the role of initial conditions. We then consider two permanent (or very persistent) changes occurred in the data after the mid-1960s: a labor productivity growth slowdown starting in 1966 as it is evident in Figure 5, Panel (a), and the deposit rate increase of the late-1970s, early-1980s as in Figure 6, Panel (a).

In particular, we impose an exogenous change in η, the probability of a project success, starting from 1966, for 15 periods. We limit the analysis to the period of more marked decline and we do not account for the partial recovery, starting in the late 1990s. Changing this structural parameter to induce a productivity growth slowdown captures the idea that the business environment deteriorated during this period, making it more difficult for entrepreneurs

[7]See Barth *et al.* (2009) for an overview of bank regulation in the United States.

to succeed. As this is a model primitive, we are agnostic on the deeper reasons for this change. The permanent nature of the change we simulate captures the idea that productivity growth stagnated hovering around 1.6-1.8% for most of the 1980s and the recovery in the early 2000s was temporary.

Next, we consider an increase in the real deposit rate for 5 periods, from 1980 to 1985. Here too, our model cannot distinguish between alternative reasons for this increase, such as the Volcker disinflation or the repeal of Regulation Q. Similarly, we consider only a 5 periods impulse. We then let the economy evolve endogenously for the remaining 30 years, from 1986 to 2015.

In the model, changing η results in an endogenous change in the path for z_{t-1}. In contrast, we can can alter r^D directly. Proposition 4 shows that the CBI decreases in aggregate labor productivity, while it increases in the cost of funds. Importantly, the number of banks correlates positively with the changes in the CBI, as suggested by Corollary 1. So, we expect that both impulses will contribute to a simultaneous increase in the CBI and the number of banks, followed by a decline. The two structural changes are phased in gradually. In the case of the change in η the economy ultimately converges to a new BGP, where g^Z is lower than in the initial one. This is because the growth rates of both the aggregate technology and the banking efficiency depend on η, with the former permanently lower and the latter is higher. In the case of the change in r^D, to economy reverts to the same BGP, which is parallel to the initial one but at a lower level due to the shock. This is because the growth rates of both the aggregate technology and the banking efficiency do not hinge on r^D. The only inter-temporal link in the agents' decision is through the aggregate state of the banking and non-financial technology levels in the model. The sequence of exogenous changes can be interpreted as a sequence of numerical comparative statics results, consistent with Proposition 4.

The two scenarios are designed consistently with the data. We keep the numerical representation of the two impulses simple and transparent, so as to evaluate the model's ability to capture the salient features of the data in a qualitative manner in two separate simulations. We will evaluate the model's ability to provide a quantitative account of the changes in the data considering a combined scenario in which we consider both structural changes at the same time.

7.2.1 The US productivity slowdown and the CBI

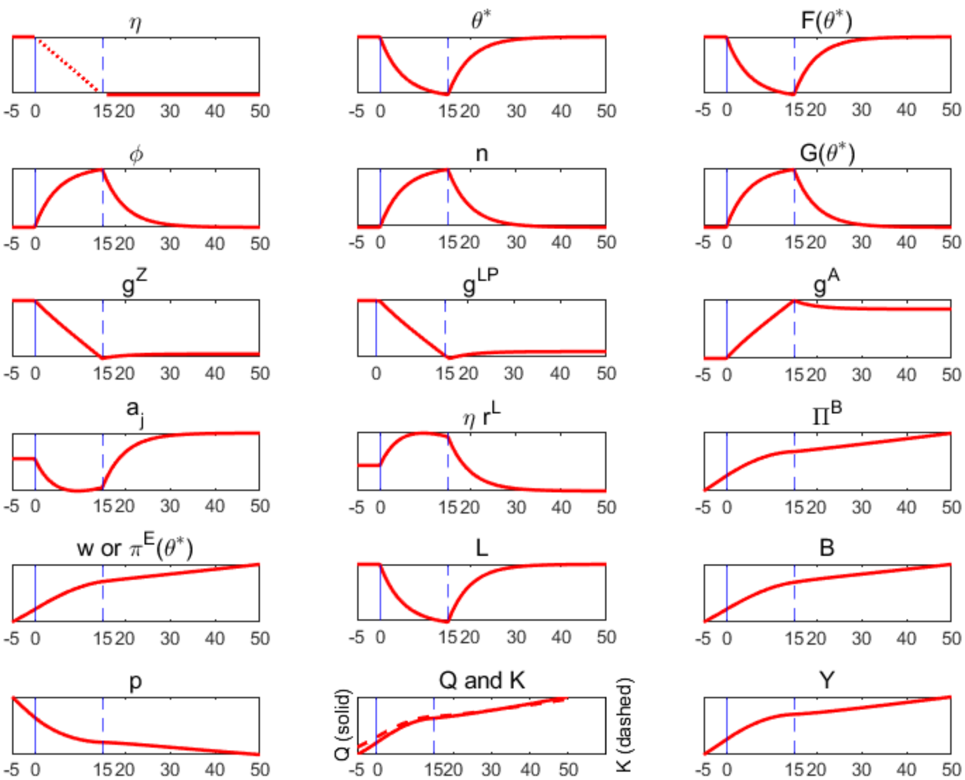
The first counterfactual simulation entails feeding to the model a labor productivity slowdown from 1966 to 1980. Labor productivity is endogenous in the model and there is no aggregate uncertainty. To induce a decline in productivity as defined in Equation (85) we adjust the entrepreneur's probability of success, η. In particular, we change η from 83% in 1965, as in the calibration above, to 79.25% in 1980, in increments of a quarter of a percentage point per year. Thereafter, η stays at the new lower level. As the only inter-temporal link between periods is through the aggregation of agent and bank specific efficiency levels, this sequence of shocks can be interpreted as a sequence of comparative statics as in Proposition 4. Once η stabilizes at its new, lower level, the economy evolves endogenously converging to a new BGP, and staying there thereafter. In the new BGP, aggregate labor productivity falls by 1.8 percentage points, to 0.3%, compared to 2.1% along the initial BGP since a lower η lowers aggregate technology growth as shown by by Equation (74). The decline is roughly consistent with the data as shown in Figure 5.

Figure 7 displays the results. A dotted line represents the exogenous change introduced. The solid and dashed lines represent the model endogenous responses. All reported variables except for a_j are aggregates, plotted in levels unless otherwise noted. During the intervention period, banks charge for a higher interest rate, r^L, to compensate for the higher default risk. Entrepreneurs' credit demand, production, and labor demand grow slower. As a result the wage growth also declines. Because the wage growth slows more than the marginal entrepreneurs' profit, more agents choose to become entrepreneurs, i.e. θ^* declines. Both the lower η and the entry into the capital good market of less skilled producers depress the growth rate of aggregate labor productivity, g^{LP}, that declines from 2.1% to 0.1% at the end of the intervention period.

A slower output growth in the capital good sector leads to a slower decline in the relative price of capital goods, p, and a slower growth in entrepreneur's profit, π^E. The slower decline is in the price of IT goods also depresses banks operating profits, discouraging banks to invest in IT equipment and making them operating at a lower level of individual efficiency, a_j. As a result, banks charge an even higher risk-adjusted rate, ηr^L.

A higher and increasing risk-adjusted lending rate implies that bankers' total profit is increasing more than than potential bankers' reservation value, $w\Gamma$, encouraging more banks to enter. More bank entry leads to a decline in the pricing power in the industry. As Equation (64)

Figure 7: A U.S. productivity slowdown



Notes: From period -5 to 0, the economy evolves along with the BGP as calibrated in Section 6.1, where $\eta = 83\%$. Starting in period 1, setting $z_0 = 1$ and $a_0 = 1$, a sequence of exogenous changes to η from period 1 through 15 is fed to the model, reducing η by a quarter of a percentage point every period. So, η goes from 83% to 79.25% in 15 declines 0f 0.25% each and then is allowed to converges to a new balanced growth path in which 79.25%. The dotted line stands for the exogenous interventions, whereas the solid and dashed lines are represented for the model outcome. The variables plotted (from top left to right) are as follows:η; the threshold value for the occupation choice solved in Equation (51); the share of workers; the CBI defined in Equation (67); the number of banks solved from Equation (58); the aggregate entrepreneurs' ability as defined in Equation (26); the gross growth rate for the aggregate technology defined in Equation (74); the gross growth rate for the aggregate labor productivity, where the labor productivity is defined in Equation (85); the gross growth rate for aggregate banking efficiency defined in Equation (77); individual bank efficiency defined in Equation (11); the interest rate charged by banks solved in Equation (21); total bank profit defined in Equation (65); the wage solved in Equation (44); the total supply of labor defined in Equation (32); and the total volume of credit intermediated by banks as defined in Equation (66); the price of capital goods as solved in Equation (43); the total amount of bank IT investment (dashed-line), defined in Equation (35) and final good producers (solid-line), solved in Equation (4); and finally the aggregate output defined in Equation (52).

shows, the number of banks affect banks' total profit in two ways: through the Lerner index and the market share of loans. However, the CBI only takes the former effect into account as the market share does not play any role when we look at the banking industry as a whole. Thus, the rise in the risk-adjusted rate dominates the decline in the Lerner index, leading to a higher CBI.

As $\tilde{\theta}^*$ declines during the intervention period, the banking sector serves a less productive pool of entrepreneurs accumulating a higher level of organizational capital. The adjustment term, $\frac{1-\mathbf{F}(\theta_t^*)}{\mathbf{F}(\theta_t^*)}$ increases leading to a higher rate of growth in aggregate bank efficiency. Note here the trade off between individual bank and industry-wide efficiency, with more slowly declining IT leading to lower individual bank efficiency, but also to a faster growing industry-wide productivity level.

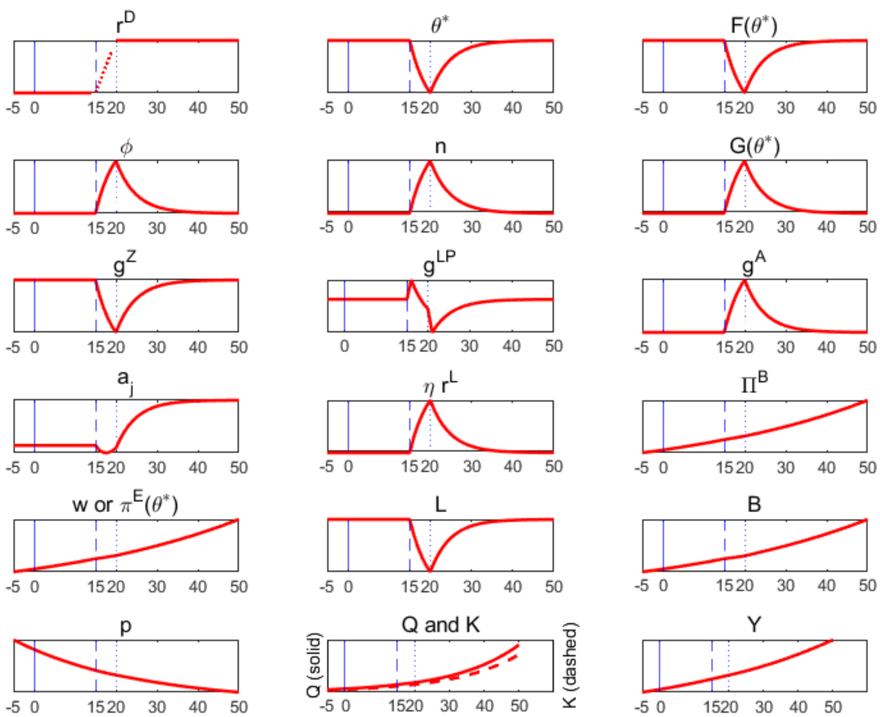
Once the the exogenous driver of the productivity slowdown ceases to perturbate the economy, the threshold level for the occupation choice is below its BGP value, $\tilde{\theta}^*$. Under Assumption (79), the gross growth rates of a_{t-1} and z_{t-1}, and $g_t^A\sqrt{g_t^Z}$, are all larger than 1. Consistent with Proposition 6, $a_t\sqrt{z_t}$ will be larger than $a_{t-1}\sqrt{z_{t-1}}$ until its gross growth rate reverts to 1. This means the threshold, will gradually increase back to $\tilde{\theta}^*$ defined in Equation (78). During the transition, as only more productive agents choose to be entrepreneurs, their average managerial ability increases. However, with a permanently lower η, the growth rate of labor productivity recovers only to 0.3%, from a trough right after the intervention of 0.1%.

As the growth rate of the aggregate technology recovers, the capital production, the final output, the total amount of loans, the bank total profit, and the wage also grow at a slower pace than the initial one, but temporarily faster than in the new BGP. The capital price resumes to decline, encouraging banks to adopt IT equipment and improve their efficiency. Thus, the risk-adjusted interest rate charged by banks declines, to a lower level than the one along with the initial BGP. Since the risk-adjusted rate is falling, the total profit by any individual bank increases less than the reservation value of potential bank entrants driven by the wage increase. As a result, banks actually exit the market and the Lerner index increases. Because the number of banks affects the industry-wide CBI only through changed pricing power, which is declining, after the intervention period, CBI declines. Also, during the transition to the new BGP, θ^* declines and banks serve a more skilled pool of borrowers. The resulting change in the relative share of workers and entrepreneurs implies that the rate of accumulation of organizational capital and the growth in aggregate bank efficiency, a_t, decline, gradually converging to a level of growth that is higher than in the initial BGP. The U.S. productivity slowdown counterfactual, therefore, implies a more concentrated and, in the aggregate, less efficient banking system in the new less productive economic environment.

7.2.2 Bank funding costs and CBI

The second counterfactual that we consider is a rise in funding costs from 1981 to 1985. Here, we permanently increase r^D by half a percentage point each period for 5 years, from 1% to 3.5%, roughly consistent with the data as shown in Figure 6. Thereafter, the economy evolves endogenously, reverting to the initial BGP with the same rate of growth aggregate productivity and banking sector technology, as the threshold for occupation choice reverts to $\tilde{\theta}^*$. Unlike the case of the productivity slowdown, the growth rate of the labor productivity eventually return to 2.1% as well, from the trough right after the intervention of 1.68%.

Figure 8: CBI and Bank Funding Costs



Notes: From period -5 to 15, the economy evolves along with the BGP as calibrated in Section 6.1, where $r^D = 1.01$. Starting in period 16, a sequence of exogenous changes to r^D are imposed, from period 16 through 20, where r^D increases by half a percentage point every period, reaching $r^D = 1.035$. Thereafter, the economy converges back to a BGP parallel to the initial one. The dotted line represents the exogenous intervention. The solid and dashed lines represent the model outcomes. See Figure 7 for the variable legend.

Figure 8 displays the results for the same variables as in Figure 7. During the intervention period, banks charge a higher interest rate, r^L, due to the higher funding cost. As the interest rate rises, banks' total profit increases relative to the entrants' reservation value, which induces more bank entry. Only the pricing power matters for CBI. As the increase in the interest rate exceeds the decline in the pricing power of banking industry (or the Lerner index), the CBI increases.

A higher interest rate implies a higher marginal cost of labor relatively to the wage, more agents choose to become entrepreneur, and the the threshold value for the occupation choice declines. This further implies a lower growth in the aggregate technology. The decline in the threshold value implies a lower average managerial ability in the pool of borrowing entrepreneurs. This means that banks accumulate more organizational capital, and the rate of growth of aggregate banking sector efficiency increases. With a lower rate of growth in aggregate technology and labor productivity, the capital goods output, the final goods output, the total volume of credit, the total profits of banks, and the wage grow at a slower rate. Meanwhile, the price of capital goods also declines at a smaller rate. This leads banks to adopt IT at a slower speed, resulting in further increase in the lending rate.

Once the shock ceases to perturbate the economy, the threshold level for the occupation choice is below $\tilde{\theta}^*$ and will gradually revert increasing to its BGP value. The average level of managerial ability across all entrepreneurs increases. Banks temporarily accumulate less organizational capital, leading to a decline in rate of growth of aggregate bank efficiency.

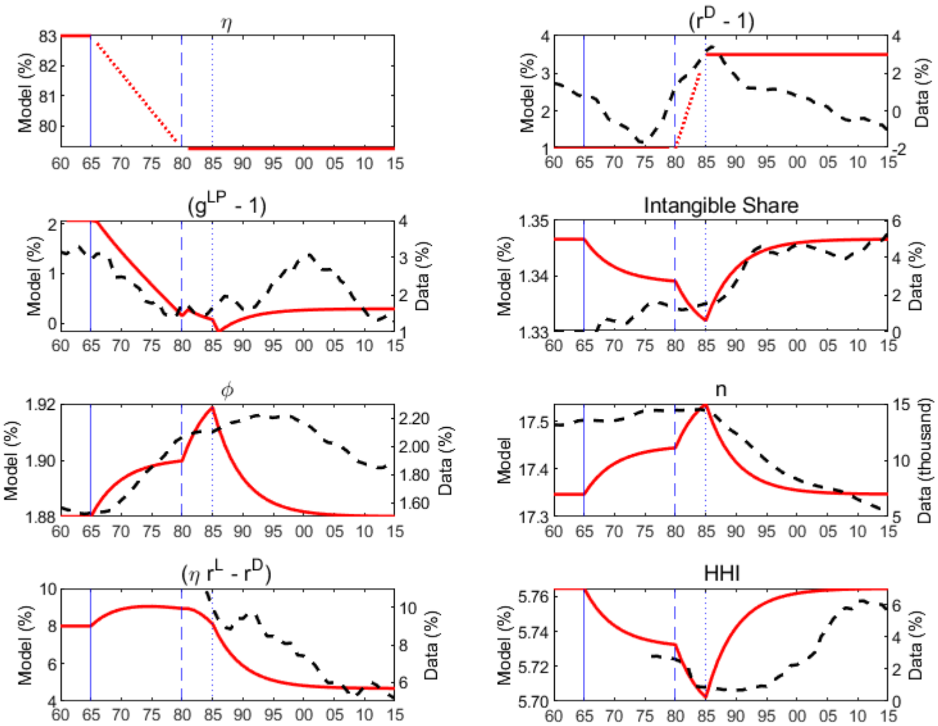
As the growth rate of the aggregate technology recovers, output, credit, bank profits, and the wage all temporarily grow faster. The pace of decline of IT equipment accelerates, encourages banks to adopt increase IT investment, improving their individual efficiency permanently. As a result, the risk-adjusted interest rate charged by banks declines, reverting to its initial BGP level, despite the permanently higher funding cost.

As the risk-adjusted rate declines, the total profit by any individual bank increases less than the reservation value, resulting in slower bank entry and higher Lerner index. Since the risk-adjusted rate decreases relatively more than the rise in the Lerner index, the CBI declines, returning to the initial level along the initial BGP.

7.3 Comparing the model with the data

In order to assess the model performance quantitatively, we combine the two experiments in one simulation and contrast the model-generated path with selected variables in the U.S. data. In particular, we assume that until 1965 the economy evolves along the balanced growth path, as calibrated in Section 6.1, where $a_{1965} = 1$, $z_{1965} = 1$, $r^D = 1.01$ and $\eta = 0.83$. Starting in 1966, the same 15-year annual decline in η discussed above is imposed, reaching $\eta = 0.7925$ in 1980. From 1981 through 1980, we impose the same sequence of changes in r^D as done before, $r^D = 1.035$ in 1985. From 1986, the economy evolves endogenously without any further exogenous change, gradually converging to a new BGP with a lower growth rate, and η and r^D at their permanently changed values.

Figure 9: Comparing the model to the data



Notes: From year 1960 to 1965, the economy evolves along the BGP as calibrated in Section 6.1. Starting in 1966, an annual decline in η of 0.25 percentage points per year is imposed for 15 years. A sequence of exogenous changes to r^D, from 1981 through 1985, follows with r^D increasing by half a percentage point every period, reaching $r^D = 1.035$ in 1985. Thereafter, the economy converges back to a new BGP. The dotted red line stands for the imposed exogenous changes. The the solid red lines represent the model implied outcomes. The dashed black line display the 10-year moving average of the corresponding data discussed above. See Figure 7 for the variable legend. In the model, we define the intangible share as the share of IT equipment over the total assets, $\frac{p_t Q_t}{\eta r_t^L B_t}$; and HH Index is equal to $\frac{1}{n}$. The vertical left axis is the scale for model variables. The vertical right axis is the scale for model variables. The dotted (red) lines represent the exogenous changes in the model. The solid solid (red) lines the model outcomes. The dashed (black) lines the data.

Figure 9 reports the results for selected variables. The left axis is the scale for model variables, while the right one is the scale for model variables. The dotted (red) lines represent the exogenous changes in the model variables, the solid (red) lines are the model outcomes, and the dashed (black) lines are the data. The figure shows that the model decline in the labor productivity from 1966 to 1980 and the rise in deposit rate between 1981 and 1985 broadly match the data through 1985. In general, the model provides a good qualitative account of the patterns in the data, even after the exogenous changes cease to drive the dynamics. The model-implied CBI, in particular, tracks its evolution in the data well from 1966 to 1985, but it drops faster than in the data thereafter. The number of banks increases less than in the data through 1985, but declines at the same speed as in the data thereafter. The model underpredicts the magnitude of the changes in the CBI, but can match more closely the number of banks.

Next, we turn to the variables that are not targeted in our calibration, investment in intangibles and bank concentration. In the model, the share of intangible assets is defined as the ratio of IT equipment over total assets, $\frac{p_t Q_t}{\eta r_t^L B_t}$, where Q_t is aggregate IT equipment, r_t^L is the interest rate, and B_t is the total volume of credit. As banks are identical and the equilibrium is symmetric, the model-based banking sector HH Index is simply $\frac{1}{n_t}$.

The model accounts remarkably well for the upward trend in the share of intangible investment after 1985, but it overpredicts the extent of the decline during the productivity slowdown. Nonetheless, we note that banks start with some IT equipment in the BGP, while in the data the series starts from from 0. The model dynamics underestimate the magnitude of the changes in the data in the case of the share of intangible investment. The model also performs well relative to the behavior of the HH Index in the data, reflecting the fact that the model can generate a change in the number of banks comparable to the one in the data, even though we clearly miss the second derivative. The model performs well in tracking the interest spread.

Overall, we conclude that the model performs well qualitatively. The quantitative performance may reflect the stylized nature of the model. Considering the role of bank regulation in determining bank entry is a possible extension to provide a more satisfactory quantitative account of the data.

8 Conclusions

The U.S. cost of financial intermediation has been relatively stable, close to 2 percent over a long period of time. During the 20th century, it has also changed significantly moving slowly at low frequency. For example, the CFI increased steadily from a trough of about 1.5% in the mod-1960s to more a peak above 2% in the early 1990s, declining thereafter.

In this paper we set up an endogenous growth model with bank intermediation whose balance growth path can match the long-run value of the CFI, showing that the banking sector market structure is the main determinant of this long-run intermediation wedge and that small changes in the CFI have large impact on long-term growth. We also zoom in the last 50 years of history and provide a model-based account of the CFI's inverted U-shape behavior, investigating the mechanisms through which economy-wide productivity changes and other structural changes in the U.S. economy might have driven these low frequency changes in the CFI.

We find that the U.S. productivity slowdown starting in the mid-1960s combined with the increase in bank funding costs due to the removal of regulation Q and the Volker disinflation can go a long way toward accounting for the CFI increase qualitatively—and to a lesser extent quantitatively. The model also matches the path of broader set of variables in the data, including particularly the increasing market concentration that accompanied the fall in the CFI starting in the early 1990s.

Exploring the role of regulation in driving the CFI in the model and an empirical investigation with bank level data of the mechanisms embedded in the model are areas of future research.

References

Aghion, P., Bergeaud, A., Boppart, T., Klenow, P. J. and Li, H. (2019). Missing Growth from Creative Destruction. *American Economic Review*, **109** (8), 2795–2822.

—, Howitt, P. and Levine, R. (2018). Financial development and innovation-led growth. In T. Beck and R. Levine (eds.), *Handbook of Finance and Development*, *Chapter 1*, Edward Elgar Publishing, pp. 3–30.

Atkeson, A. and Kehoe, P. J. (2005). Modeling and measuring organization capital. *Journal of political Economy*, **113** (5), 1026–1053.

Axtell, R. L. (2001). Zipf Distribution of U.S. Firm Sizes. *Science*, **293** (5536), 1818–1820.

Barth, J. R., Li, T. and Lu, W. (2009). Bank Regulation in the United States. *CESifo Economic Studies*, **56** (1), 112–140.

Bazot, G. (2017). Financial Consumption and the Cost of Finance: Measuring Financial Efficiency in Europe (1950–2007). *Journal of the European Economic Association*, **16** (1), 123–160.

Bernanke, B., Gertler, M. and Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. In J. B. Taylor and M. Woodford (eds.), *Handbook of Macroeconomics*, vol. 1C, *Chapter 21*, Elsevier, pp. 1341–1393.

Cetorelli, N., Hirtle, B., Morgan, D. P., Peristiani, S. and Santos, J. A. C. (2007). Trends in Financial Market Concentration and their Implications for Market Stability. *Economic Policy Review*, **13** (1), 33–51.

Corbae, D. and Levine, R. (2020). Competition, Stability, and Efficiencyin the Banking Industry, university of Wisconsi.

Drechsler, I., Savov, A. and Schnabl, P. (2017). The Deposits Channel of Monetary Policy. *The Quarterly Journal of Economics*, **132** (4), 1819–1876.

—, — and — (2020). The Financial Origins of the Rise and Fall of American Inflation, manuscript.

Fernald, J. G. (2012). *A Quarterly, Utilization-Adjusted Series on Total Factor Productivity.* Working Paper 2012-19, Federal Reserve Bank of San Francisco, San Francisco.

Gabaix, X. (2016). Power Laws in Economics: An Introduction. *Journal of Economic Perspectives*, **30** (1), 185–206.

Gersbach, H., Rochet, J.-C. and Scheffel, M. (2019). *Financial Intermediation, Capital Accumulation and Crisis Recovery.* Research Paper 19-62, Swiss Finance Institute.

Greenwood, J., Sanchez, J. M. and Wang, C. (2010). Financing Development: The Role of Information Costs. *American Economic Review*, **100** (4), 1875–1891.

Greenwood, R. and Scharfstein, D. (2013). The Growth of Finance. *Journal of Economic Perspectives*, **27** (2), 3–28.

Jayaratne, J. and Strahan, P. E. (1996). The Finance-Growth Nexus: Evidence from Bank Branch Deregulation. *The Quarterly Journal of Economics*, **111** (3), 639–670.

King, R. G. and Levine, R. (1993). Finance and Growth: Schumpeter Might Be Right. *The Quarterly Journal of Economics*, **108** (3), 717–737.

Levine, R., Loayza, N. and Beck, T. (2000). Financial intermediation and growth: Causality and causes. *Journal of Monetary Economics*, **46** (1), 31–77.

McCord, R., Prescott, E. S. and Sablik, T. (2015). *Explaining the Decline in the Number of Bankssince the Great Recession.* Economic Brief EB15-03, Federal Reserve Bank of Richmond, Richmond.

McGrattan, E. R. and Prescott, E. C. (2005). Taxes, regulations, and the value of us and uk corporations. *The Review of Economic Studies*, **72** (3), 767–796.

Melitz, M. J. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica*, **71** (6), 1695–1725.

Philippon, T. (2015). Has the US Finance Industry Become Less Efficient? On the Theory and Measurement of Financial Intermediation. *American Economic Review*, **105** (4), 1408–1438.

— (2016). *The FinTech Opportunity.* NBER Working Paper 22476, National Bureau of Economic Research.

— and Reshef, A. (2013). An International Look at the Growth of Modern Finance. *Journal of Economic Perspectives*, **27** (2), 73–96.

Pierri, N. and Timmer, T. (2020). The Importance of Technology in Banking during a Crisis, unpublished.

Rajan, R. G. and Zingales, L. (1998). Financial Dependence and Growth. *The American Economic Review*, **88** (3), 559–586.

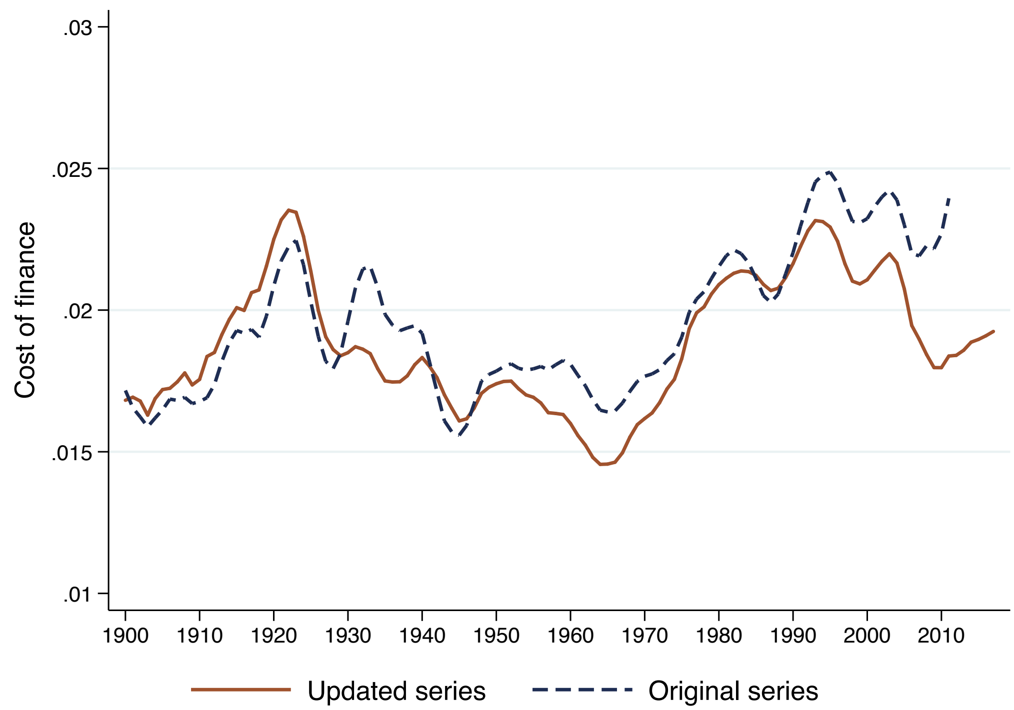
Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, **21** (2), 265–293.

Williamson, S. D. (1986). Costly monitoring, financial intermediation, and equilibrium credit rationing. *Journal of Monetary Economics*, **18** (2), 159–179.

A Updated Estimate of the Cost Of Financial Intermediation

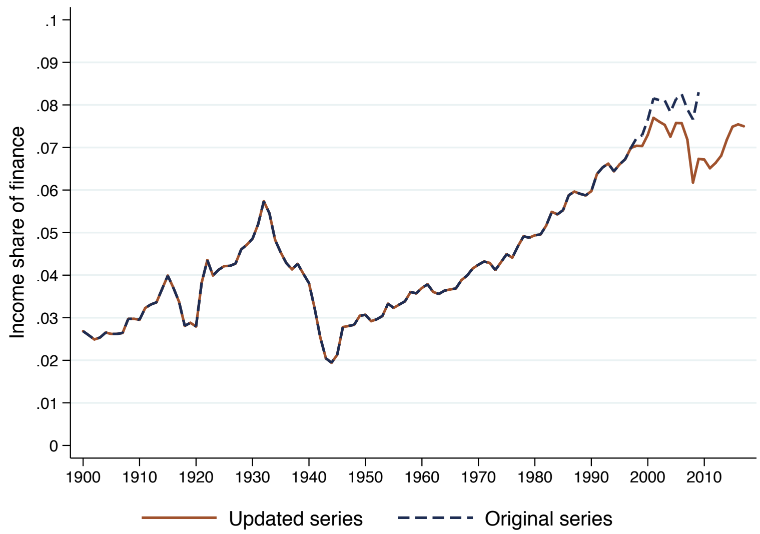
The unit cost of finance is the ratio of finance income to intermediated assets as estimated by Philippon (2015). The main source if the Flow of Funds of the BEA. A comparison between the original and the updated series is reported below, for the overall cost as well as its components, the value added and the output of the finance industry. Both series are expressed as a share of GDP. Finance income is the domestic income of the finance and insurance industries, i.e., aggregate income minus net exports. Intermediated assets include debt and equity issued by non-financial firms, household debt, and various assets providing liquidity service.

Figure A.1: Cost of Financial Intermediation
Original and Updated Series



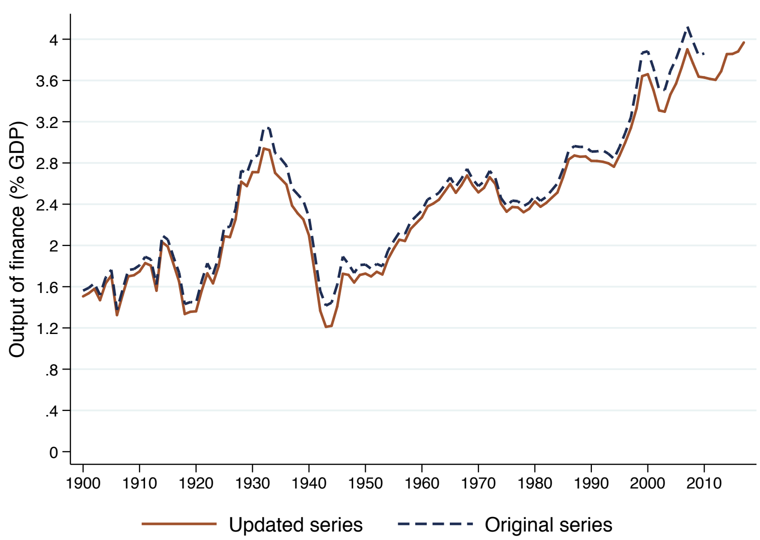
Notes: The figure shows the unit cost of financial intermediation. The dashed blue line is the original series as in Philippon (2015), while the solid maroon line is the updated series.

Figure A.2: Value Added of the Finance Industry
Original and Updated Series



Notes: The figure shows the unit cost of financial intermediation. The dashed blue line is the original series as in Philippon (2015), while the solid maroon line is the updated series.

Figure A.3: Output of the Finance Industry:
Original and Updated Series



Notes: The figure shows the unit cost of financial intermediation. The dashed blue line is the original series as in Philippon (2015), while the solid maroon line is the updated series.

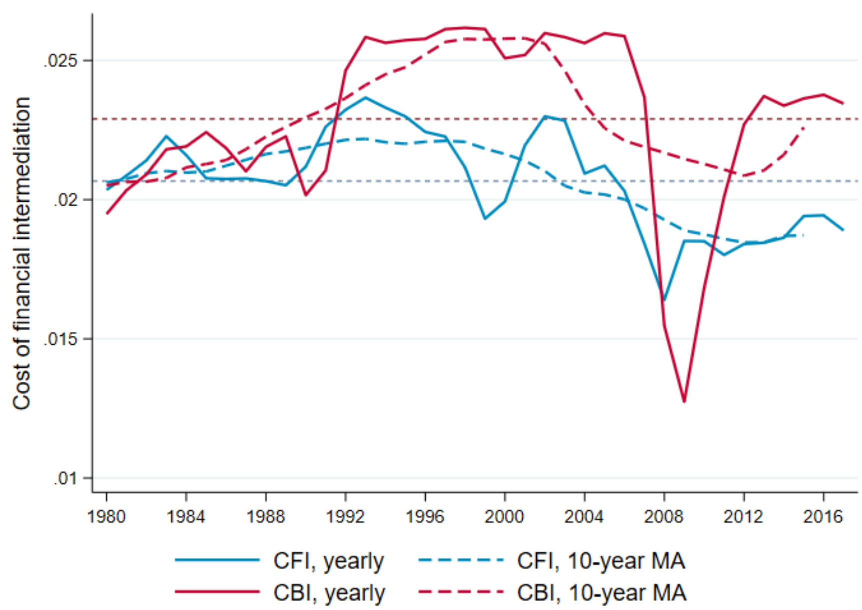
B The Cost of Intermediation in Finance and in Banking

Our measure of the CBI is the ratio of banking value added over bank intermediated assets, which we compute from Call Reports for U.S. commercial banks from 1976 to 2017:

$$CBI = \frac{Salaries + Net\ Income}{Cash + Loans + Securities + Equity}. \tag{A1}$$

As we noted earlier, the U.S. CFI, as measured from U.S. Flows of Funds Data, exhibits significant medium term fluctuations around a 2% long-run average, dropping from 2.5 percent in the early 1920s, to less than 1.5 percent in 1965, to climb back to 2.4 percent in the mid 1990s, and then revert toward its long-term average. These are large swings relative to the long-term average, as small variations are likely to have a relatively large impact on aggregate borrowing. As Figure B.1 shows, the CBI tracks closely the overall CFI during the period over which they overlap, with a correlation of 0.72.

Figure B.1: The evolution of the CFI and CBI (1976-2017)



Notes: The figure plots the cost of financial intermediation (CFI), as computed by updating the series by Philippon (2015)—see also Figure 1—and the cost of bank intermediation (CBI) based on Flow of Funds Data as in Equation (A1). The solid lines plot the yearly data, while the dotted lines are 10-year moving averages. The dashed horizontal lines are the long-term averages. The correlation coefficient between the yearly series of CFI and CBI is equal to 0.72.

C Proofs of Propositions and Corollary

In this Appendix, we provide the proofs of the lemmas and propositions stated in the main text.

C.1 Proof of Proposition 1

We first solve the optimal problem for banks in this economy, and then show that there exists a unique solution to the banks' problem. The associated Lagrangian is

$$\begin{aligned}\mathcal{L} &= \eta r^L_{jt}(\theta) b_{jt}(\theta) - \frac{r^D b_{jt}(\theta)}{a_{jt}} - (1-\eta)\frac{e_{jt}(\theta) r^D b_{jt}(\theta)}{a_{jt}} \\ &+\lambda_1 \left(\frac{p_t \theta z_{t-1}}{w_t^{\xi}} \left(\sum_{i=1}^{n_t} b_{it}(\theta) \right)^{\xi-1} - r^{\theta,0}_{jt}(\theta) \right) b_{jt}(\theta) \\ &+\lambda_2 \left(\mathbf{Pr}\left(e_{jt}(\theta)\right) r^{\theta,0}_{jt}(\theta) - r^L_{jt}(\theta) \right) b_{jt}(\theta) \\ &+\lambda_3 \left(\frac{\xi p_t \theta z_{t-1}}{w_t^{\xi}} \left(\sum_{i=1}^{n_t} b_{it}(\theta) \right)^{\xi-1} - r^L_{jt}(\theta) \right) b_{jt}(\theta)\end{aligned}$$

The first order conditions are

$$(r^L_{jt}(\theta)) \qquad \eta = \lambda_2 + \lambda_3 \tag{A1}$$

$$(r^{\theta,0}_{jt}(\theta)) \qquad \lambda_1 = \frac{e_{jt}(\theta)\lambda_2}{e_{jt}(\theta)+\sigma} \tag{A2}$$

$$(e_{jt}(\theta)) \qquad (1-\eta)\frac{r^D}{a_{jt}} = \frac{\sigma r^{\theta,0}_{jt}(\theta)\lambda_2}{(e_{jt}(\theta)+\sigma)^2} \tag{A3}$$

$$(b_{jt}(\theta)) \qquad \eta r^L_{jt}(\theta) - (1+(1-\eta)e_{jt}(\theta))\frac{r^D}{a_{jt}} = \frac{\left(\lambda_1 + \xi\lambda_3\right) p_t \theta z_{t-1}}{w_t^{\xi}\left(\sum_{j=1}^{n} b_{jt}\right)^{1-\xi}} \frac{(1-\xi) b_{jt}}{\sum_{j=1}^{n} b_{jt}} \tag{A4}$$

$$(\lambda_1) \qquad \lambda_1 \left(\frac{p_t \theta z_{t-1}}{w_t^{\xi}} \left(\sum_{i=1}^{n_t} b_{it}(\theta) \right)^{\xi-1} - r^{\theta,0}_{jt}(\theta) \right) = 0 \tag{A5}$$

$$(\lambda_2) \qquad \lambda_2 \left(\mathbf{Pr}\left(e_{jt}(\theta)\right) r^{\theta,0}_{jt}(\theta) - r^L_{jt}(\theta) \right) = 0 \tag{A6}$$

$$(\lambda_3) \qquad \lambda_3 \left(\frac{\xi p_t \theta z_{t-1}}{w_t^{\xi}} \left(\sum_{i=1}^{n_t} b_{it}(\theta) \right)^{\xi-1} - r^L_{jt}(\theta) \right) = 0 \tag{A7}$$

Note that $\lambda_3 > 0$ due to Equation (9). Using Equation (A3), we have $\lambda_2 > 0$. Combining this with Equation (A2), we have $\lambda_1 > 0$. Thus we have

$$r^{\theta,0}_{jt}(\theta) = \frac{p_t \theta z_{t-1}}{w_t^{\xi}} \left(\sum_{i=1}^{n_t} b_{it}(\theta) \right)^{\xi-1} \tag{A8}$$

$$\mathbf{Pr}\left(e_{jt}(\theta)\right) r^{\theta,0}_{jt}(\theta) = r^L_{jt}(\theta) \tag{A9}$$

Substituting $r_{jt}^{\theta,0}(\theta)$ from Equation (A8) and $r_{jt}^{L}(\theta)$ from Equation (A8) into Equation (A9), we have $\mathbf{Pr}\left(e_{jt}(\theta)\right) = \frac{e_{jt}(\theta)}{e_{jt}(\theta)+\sigma} = \xi$. This implies

$$e_{jt} = \frac{\xi\sigma}{1-\xi} \tag{A10}$$

Using Equation (A8), we have $\lambda_2 = \frac{\lambda_1}{\xi}$, which, combining with Equation (A1), implies

$$\frac{\lambda_1}{\xi} + \lambda_3 = \eta \tag{A11}$$

Plugging Equation (A11) and (9) into Equation (A4), we have, under symmetry,

$$\eta r_{jt}^{L}(\theta) = \frac{\left(1+\tilde{\sigma}\right) r^D}{\left(1 - \frac{1-\xi}{n_t}\right) a_{jt}} \tag{A12}$$

where $\tilde{\sigma} = \frac{(1-\eta)\xi}{1-\xi}\sigma$. Plugging Equation (A12) into (9), we have

$$b_t(\theta) = \left(\frac{\xi p_t \theta z_{t-1}}{w_t^{\xi}} \frac{\eta\left(1 - \frac{1-\xi}{n_t}\right) a_{jt}}{\left(1+\tilde{\sigma}\right) r^D} \right)^{\frac{1}{1-\xi}} \tag{A13}$$

C.2 Proof of Proposition 2

Using Equation (48), for type–θ agent, the value of being entrepreneur relative to being a worker is

$$\begin{aligned} \frac{\pi^E\left(\theta, \hat{\theta}_t\right)}{w_t\left(\hat{\theta}_t\right)} &= \left(\frac{1}{\xi} - 1\right) \mathbf{F}\left(\hat{\theta}_t\right) \eta r_t^L\left(\hat{\theta}_t\right) \frac{\theta^{\frac{1}{1-\xi}}}{G\left(\hat{\theta}_t\right)} \qquad \text{(A14)} \\ &= \frac{2(1-\xi)}{(2-3\xi)\xi} \left(1 - \xi + \frac{2-3\xi}{2} \frac{M}{\Gamma} \frac{\left(1+\tilde{\sigma}\right) r^D \left(\mathbf{F}\left(\hat{\theta}_t\right)\right)^{1-\frac{\xi}{2}}}{\xi\kappa a_{t-1} \left(\frac{\eta}{2} M z_{t-1} \left(G\left(\hat{\theta}_t\right)\right)^{1-\xi}\right)^{\frac{1}{2}}} \right)^2 \frac{\Gamma}{M} \frac{\theta^{\frac{1}{1-\xi}}}{G\left(\hat{\theta}_t\right)}, \end{aligned}$$

where the second equality follows from Equation (41) and (47).

Clearly, the relative profit of marginal entrepreneur to worker, namely $\frac{\pi^E\left(\hat{\theta}_t, \hat{\theta}_t\right)}{w_t\left(\hat{\theta}_t\right)}$ is increasing in $\hat{\theta}_t$, because $\mathbf{F}\left(\hat{\theta}_t\right)$ and $G\left(\hat{\theta}_t\right)$ respectively increases and decreases in $\hat{\theta}_t$. Moreover, we have

$$\frac{\pi^E\left(\underline{\theta}, \underline{\theta}\right)}{w_t\left(\underline{\theta}\right)} = \frac{2(1-\xi)^3}{(2-3\xi)\xi} \frac{\Gamma}{M} \frac{\underline{\theta}^{\frac{1}{1-\xi}}}{G\left(\underline{\theta}\right)} < 1 \tag{A15}$$

where the inequality is ensured by the Assumption (49). $\lim_{\hat{\theta}_t \to \infty} \frac{\pi^E\left(\hat{\theta}_t, \hat{\theta}_t\right)}{w_t\left(\hat{\theta}_t\right)} = \infty$. Therefore, there must exist a $\tilde{\theta}^*$ where $\frac{\pi^E\left(\theta^*, \theta^*\right)}{w_t\left(\hat{\theta}_t\right)} = 1$. Moreover, the uniqueness is guaranteed by the monotonicity of $\frac{\pi^E(\hat{\theta}_t,\hat{\theta}_t)}{w_t(\hat{\theta}_t)}$.

C.3 Proof of Proposition 3

As shown in Section C.2, the left hand side (LHS) of Equation (51) is increasing in θ_t^*. Also, it is straightforward to show that the LHS is decreasing in z_{t-1}, a_{t-1}, and η, and increasing in r^D. Taking the total differentiation of Equation (51) respectively against z_{t-1}, a_{t-1}, and η, we have

$$\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ \theta_t^*}\frac{\mathbf{d}\ \theta_t^*}{\mathbf{d}\ z_{t-1}}+\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ z_{t-1}}=0 \tag{A16}$$

$$\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ \theta_t^*}\frac{\mathbf{d}\ \theta_t^*}{\mathbf{d}\ a_{t-1}}+\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ a_{t-1}}=0 \tag{A17}$$

$$\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ \theta_t^*}\frac{\mathbf{d}\ \theta_t^*}{\mathbf{d}\ r^D}+\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ r^D}=0 \tag{A18}$$

$$\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ \theta_t^*}\frac{\mathbf{d}\ \theta_t^*}{\mathbf{d}\ \eta}+\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ \eta}=0 \tag{A19}$$

Using Equation , we have

$$\frac{\mathbf{d}\ \theta_t^*}{\mathbf{d}\ z_{t-1}}=-\frac{\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ \theta_t^*}}{\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ z_{t-1}}}>0 \tag{A20}$$

$$\frac{\mathbf{d}\ \theta_t^*}{\mathbf{d}\ a_{t-1}}=-\frac{\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ \theta_t^*}}{\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ a_{t-1}}}>0 \tag{A21}$$

$$\frac{\mathbf{d}\ \theta_t^*}{\mathbf{d}\ r^D}=-\frac{\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ \theta_t^*}}{\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ r^D}}>0 \tag{A22}$$

$$\frac{\mathbf{d}\ \theta_t^*}{\mathbf{d}\ \eta}=-\frac{\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ \theta_t^*}}{\frac{\partial\ LHS\ of\ Eq.\ (51)}{\partial\ \eta}}>0 \tag{A23}$$

C.4 Proof of Collary 1

Using Equation (58), we show that $\frac{\partial n_t}{\partial \theta_t^*}<0$. Using proposition 3, we have

$$\frac{\mathbf{d}\ n_t}{\mathbf{d}\ z_{t-1}}<0 \tag{A24}$$

$$\frac{\mathbf{d}\ n_t}{\mathbf{d}\ a_{t-1}}<0 \tag{A25}$$

$$\frac{\mathbf{d}\ n_t}{\mathbf{d}\ r^D}>0 \tag{A26}$$

$$\frac{\mathbf{d}\ n_t}{\mathbf{d}\ \eta}<0 \tag{A27}$$

C.5 Proof of Proposition 4

Using Equation (68), we show that $\frac{\partial \phi_t}{\partial \theta_t^*} < 0$. Using proposition 3, we have

$$\frac{\mathbf{d}\ \phi_t}{\mathbf{d}\ z_{t-1}} < 0 \tag{A28}$$

$$\frac{\mathbf{d}\ \phi_t}{\mathbf{d}\ a_{t-1}} < 0 \tag{A29}$$

$$\frac{\mathbf{d}\ \phi_t}{\mathbf{d}\ r^D} > 0 \tag{A30}$$

$$\frac{\mathbf{d}\ \phi_t}{\mathbf{d}\ \eta} < 0 \tag{A31}$$

C.6 Proof of Proposition 5

Define function $h(\theta)$ as

$$h(\theta) = \tau\kappa \left(\frac{(1-\xi)^{1-\xi}\xi\psi}{(2-3\xi)(\psi-1)} \left(1 - \xi - \frac{1}{\psi}\right)^{\xi} \Gamma \right)^{\frac{1}{2}} \tilde{\theta}^* \left(\left(\frac{\tilde{\theta}^*}{\underline{\theta}} \right)^{\psi} - 1 \right)^{\frac{\xi}{2}-\delta} \tag{A32}$$

Under Assumption (79), we have $h'(\theta) < 0$ with $\lim_{\theta \to \underline{\theta}} h(\theta) = \infty$ and $\lim_{\theta \to \infty} h(\theta) = 0$. Therefore, there exists a unique $\tilde{\theta}^*$ such that $h\left(\tilde{\theta}^*\right) = 1$. For any $\{\tilde{z}_0, \tilde{a}_0\}$ satisfying Equation (80), then we have $g^A\left(\tilde{\theta}^*\right)\sqrt{g^Z\left(\tilde{\theta}^*\right)} = 1$. Thus $a_t\sqrt{z_t} = a_{t-1}\sqrt{z_{t-1}}$ for any $t \geq 1$. Using Equation (51), we have $\theta_{t+1}^* = \theta_t^* = \tilde{\theta}^*$. Equation (58), (46), (75), and (68) respectively implies that n, ηR^L, a_j, and ϕ all stay constant.

C.7 Proof of Proposition 6

Along the BGP, Equation (51) implies that $g_t^A\left(\tilde{\theta}^*\right)\left(g_t^Z\left(\tilde{\theta}^*\right)\right)^{\frac{1}{2}} = 1$ with a fixed threshold level for the occupation choice. Thus, if $\theta_t^* < \tilde{\theta}^*$, $g_t^A\left(\theta_t^*\right)\left(g_t^Z\left(\theta_t^*\right)\right)^{\frac{1}{2}} > 1$, which implies that $a_t\sqrt{z_t} > a_{t-1}\sqrt{z_{t-1}}$, and using the proof of Section C.3, we have $\theta_{t+1}^* > \theta_t^*$. Otherwise, $g_t^A\left(\theta_t^*\right)\left(g_t^Z\left(\theta_t^*\right)\right)^{\frac{1}{2}} < 1$, implying that $a_t\sqrt{z_t} < a_{t-1}\sqrt{z_{t-1}}$, and thus $\theta_{t+1}^* < \theta_t^*$