Macro Shocks and Firm Dynamics with Oligopolistic Financial Intermediaries

Alessandro T. Villa*

June 25, 2026

Abstract

Motivated by a secular increase in the concentration of the U.S. banking industry, I develop a new macroeconomic model with oligopolistic financial intermediaries and heterogeneous firms. Market power allows banks to price discriminate and charge firm-specific markups, exerting greater market power over productive and more financially constrained firms. This dampens capital accumulation and amplifies the effects of macroeconomic shocks. During a crisis, banks exploit the higher share of financially constrained firms to extract higher markups, inducing a larger decline in real activity. When a large bank fails, the remaining banks use their increased market power to restrict credit supply, worsening and prolonging the downturn.

JEL Codes: D43, E44, G12, G21, L11.

Keywords: Dynamic Financial Oligopoly, Endogenous Financial Markups, Heterogeneous Firms, Firm Dynamics, Micro-Founded Financial Frictions, Price Discrimination.

---

*Federal Reserve Bank of Chicago. Mailing address: 230 S. La Salle St., Chicago, IL 60604, USA. Email: alessandro.tenzin.villa@gmail.com. Acknowledgments: I am indebted to my advisors Andrea Lanteri, Lukas Schmid, Giuseppe Lopomo, Francesco Bianchi, and Pietro Peretto for their support and advice. I would also like to thank Joel David, François Gourio, Leonardo Melosi, Pablo D'Erasmo, Nicola Persico, Stijn G. Van Nieuwerburgh, Dean Corbae, Vincenzo Quadrini, Robert Barsky, Virgiliu Midrigan, Roberto Steri, Fabio Schiantarelli, Matthias Kehrig and Cosmin L. Ilut for their helpful comments and encouragement. I benefited from fruitful conversations with numerous seminar participants at Computing in Economics and Finance 2019, BFI Macro Finance Research Program 2020, Econometric Society/Bocconi World Congress 2020, University of Surrey, Duke University, University of Essex, Boston College, Chicago Fed, University of Washington, Federal Reserve Board, Penn State University, and University of Wisconsin-Madison. Disclaimer: The views expressed in this paper do not represent the views of the Chicago Fed or the Federal Reserve System. I declare that I have no relevant or material financial interests that relate to the research described in the paper. All errors are my own.

1 Introduction

The banking industry has become increasingly concentrated over the past four decades, with the asset-market share of the four largest U.S. banks rising from 10% in 1984 to about 40% in 2022. Moreover, the Lerner index – a metric of market power in the banking industry – increased from 0.17 in 1984 to 0.48 in 2022, pointing to a sizable increase in markups.[1] A large and influential literature has studied the interactions between financial markets, firm dynamics, and aggregate dynamics but has typically assumed perfectly competitive financial intermediaries; thus, it does not speak to this trend of increasing banking sector consolidation and markups.[2] Moreover, extensive empirical evidence suggests that bank market power affects younger and older firms differently.[3]

In this paper, I study the role of imperfect competition in the financial intermediation sector for firm investment and financing dynamics, as well as for the transmission of macroeconomic shocks. First, I develop a dynamic general equilibrium model that incorporates an oligopolistic financial sector (including entry and exit decisions) with heterogeneous firms. The framework formalizes the notion that bank market power has heterogeneous effects across firms with different characteristics, such as size, age, or productivity. Imperfect competition enables financial intermediaries to charge firm-specific markups that depend on the idiosyncratic characteristics of the firms to which they lend. In particular, banks exert a higher degree of market power on firms that are more financially constrained and have a high marginal productivity of capital. These firms have worse outside options (e.g., a high cost of non-bank finance); hence, they exhibit a higher and less elastic demand for credit. This mechanism creates endogenous financial frictions that slow the growth of firms that operate in more concentrated credit markets. Second, I show that although bank market power persistently dampens capital accumulation, its effects intensify during downturns—when firm credit needs are greatest—thereby amplifying the transmission of macroeconomic shocks. During a crisis, oligopolistic banks exploit the larger number of financially constrained firms to extract higher markups, inducing a larger misallocation of credit (hence, capital) and a larger decline in real activity. Notably, since my model features non-atomistic banks, I can study market structure changes in the intermediation sector (e.g., the failure of a large intermediary and the entry of a new bank). When a single big bank fails, surviving banks use their increased market power to restrict credit supply, which amplifies and prolongs the recession. The results suggest that bank market power should be an important source of concern for policymakers deciding whether to bail out a large intermediary.

[1] See the left panel of Figure 1 for bank concentration and the right panel for the Lerner index.

[2] See, for instance, Bernanke and Gertler (1989), Covas and den Haan (2011), Jermann and Quadrini (2012), Khan and Thomas (2013), and Midrigan and Xu (2014).

[3] See, for example, Petersen and Rajan (1995), Rajan and Zingales (1998), Cetorelli and Gambera (2001), Black and Strahan (2002), and Cetorelli and Strahan (2006). Firms reliant on external funding via bank loans, e.g. small and private firms, can become financially constrained when credit conditions tighten (Holmstrom and Tirole, 1997; Diamond and Rajan, 2005; and Chodorow-Reich, 2013). See Subsection 5.2.1 for details.

In summary, the paper makes three contributions. First, to the best of my knowledge, I am the first to develop a macroeconomic model that incorporates oligopolistic banks and heterogeneous firms that formalizes the idea that bank market power has different effects across firms with different characteristics. The model reveals a mechanism of endogenous financial frictions through which bank competition can play a role in shaping the speed at which firms grow, impacting aggregate productivity and output. More precisely, limited competition enables banks to price discriminate and charge firm-specific markups, exerting a higher degree of market power on firms with a high marginal productivity of capital that are more financially constrained. The resulting dispersion of markups induces credit misallocation and thus capital misallocation, reducing aggregate productivity. Second, I examine how bank market power shapes the transmission of aggregate shocks and find that a marginally more restricted credit supply during periods of elevated demand dampens capital accumulation, thereby amplifying the effects of these shocks. Third, I make a methodological contribution by extending existing heterogeneous firms algorithms to solve for the stationary equilibrium and transitional dynamics in the presence of a continuum of non-competitive markets and strategic interactions by exploiting generalized Euler equations.[4]

Succinctly, the model works as follows. Firms make optimal capital structure decisions by balancing equity and debt financing, generating an endogenous dynamic demand for loans. Banks are large (i.e., non-zero mass) players and firms are a continuum of followers in a Stackelberg fashion (i.e., each financial intermediary takes dynamic loan demand by firms as given and competes to supply funding to each individual firm). Intermediaries make strategic decisions by internalizing the effects of their actions on current and future bank behavior, firm decisions, and the aggregate economy. In such an environment, the bank's optimal equilibrium choice of loan supply is determined by solving the aforementioned generalized Euler equations. Each generalized Euler equation is a standard Euler equation, except that it contains a firm-specific elasticity that measures the sensitivity of the future interest rate with respect to the current supply of loans. The model generates firm-specific credit spreads that accrue to banks, which include default premia and markups.

The analysis proceeds in two main steps. First, I develop a stylized two-period model that I use to derive analytical insights on the role of oligopolistic intermediaries for firm dynamics and aggregate outcomes. Second, I build an infinite-horizon version of the model and discuss the impact of the main economic mechanism in greater detail. I use the model to show that the time-varying cross-sectional effects of financial markups play a significant role in amplifying the impacts of macroeconomic shocks. I calibrate the model to match several financial and macroeconomic variables using Federal Deposit Insurance Corporation (FDIC) data. Compared to a perfectly competitive setting, the calibrated model incorporating both bank market power and bank default predicts an amplification of aggregate credit spreads

[4]Note that the optimal fiscal policy literature uses generalized Euler equations and Markov perfect equilibria in macroeconomics (Klein and Ríos-Rull, 2003; Klein, Krusell, and Ríos-Rull, 2008; and Clymo and Lanteri, 2020).

of 25 basis points at the peak of the Great Recession, resulting in a 0.5 percentage point decline in output. In the long term, the permanent loss of a major intermediary raises credit spreads by 15 basis points, leading to a sustained 0.2 percentage point decline in output.

In particular, I use the model to investigate the role that bank market power plays in the transmission of three unexpected aggregate shocks: (i) a decrease in aggregate Total Factor Productivity (TFP) calibrated to the size of the Great Recession and, by itself, not sufficiently large to induce a bank failure, (ii) a temporary change in bank market structure (in the model, the first shock combined with an idiosyncratic shock to the assets of the big banks calibrated to induce one bank to default with a subsequent new bank entry), and (iii) the second shock combined with a permanent increase in the fixed cost of entry into the credit market (that induces a permanent change to the bank market structure in the spirit of capturing the long-run trend of consolidation in the banking sector). I find that in each of these cases, bank market power plays a significant role in shaping—and, in particular, amplifying—the economy's response to the exogenous shock.

A decrease in aggregate TFP, with an associated increase in the aggregate probability of firm default, induces a higher proportion of young, more financially constrained firms with a high marginal productivity of capital. In these conditions, a more concentrated banking sector can control the supply of credit more tightly by extracting higher markups from these firms, leading to higher interest rates. This mechanism allows banks to compensate for the larger losses due to defaults, but it leads to a larger decline in real activity, amplifying the recession. When this shock is also combined with a lower firm entry rate, then imperfect competition in the financial intermediation sector leads to a larger and delayed amplification effect at the peak that fades away as new firms enter the production sector.[5]

When one large financial intermediary fails, the change in market structure lowers the supply of credit to firms, slowing down the economy. The surviving banks extend more credit to firms in order to capture the market share of the defaulted bank. However, the speed of this adjustment is dampened by the decreased level of competition. As a result of both credit constraints and market power, the aggregate volume of credit drops sharply in the short run. In the long run, I analyze two different scenarios. First, I analyze a scenario where one bank reenters the credit market as the financial and economic conditions ease. Second, I analyze a scenario where no bank reenters (in the model, this is obtained through a permanent increase in the fixed cost of entry into the credit market). In the former case, it takes several quarters for the economic conditions to ease sufficiently so that it is profitable for one bank to reenter. This persistently depresses investment, TFP, and output. In the latter case, the resulting increase in bank market power further amplifies and prolongs the recession and, in the long run, the economy stabilizes at a lower level of total credit, which results in permanently lower investment, output, and productivity. The result suggests that bank market power may be an important source of concern for policymakers deciding whether to bail out a big bank.

[5]For instance, during the Great Recession we observed a decline of the firm entry rate.

Related Literature. This paper contributes to two main strands of the literature: (i) firm dynamics under financial frictions, and (ii) macroeconomic models with imperfectly competitive financial intermediaries. A key contribution is to integrate these strands by studying how bank market power affects firm-level investment and financing behavior in a heterogeneous firm environment. While models of firm dynamics often assume competitive credit markets, and macro-finance models with bank competition typically abstract from firm heterogeneity, I bridge this gap by embedding a dynamic financial oligopoly in a heterogeneous firm framework.

Firm Dynamics and Credit Markets. A large literature studies firm dynamics under borrowing constraints or financial frictions (e.g., Gomes, 2001; Cooley and Quadrini, 2001; Hennessy and Whited, 2005, 2007; Jermann and Quadrini, 2012; Khan and Thomas, 2013). However, these models typically assume competitive credit markets. This paper departs from that assumption by introducing imperfectly competitive banks that set cross-sectional financial markups, thereby extending standard macroeconomic frameworks of firm dynamics to include an endogenous, imperfectly competitive credit supply mechanism.

Strategic interaction among banks generates endogenous dispersion in borrowing costs, which in turn affects capital allocation. This mechanism connects to recent work emphasizing dispersion and misallocation (e.g., Lanteri, 2018; David, Schmid, and Zeke, 2022) and aligns with empirical evidence showing that credit market structure can shape firm-level outcomes (e.g., Petersen and Rajan, 1995; Rajan and Zingales, 1998; Black and Strahan, 2002; Cetorelli and Gambera, 2001; Cetorelli and Strahan, 2006).

Macroeconomics with Financial Intermediaries. Several papers analyze the role of financial intermediaries in macroeconomics, either by focusing on imperfect competition among banks (e.g., Corbae and D'Erasmo, 2021) or by examining the interaction between credit constraints and the financial intermediation sector (e.g., Elenev, Landvoigt, and Van Nieuwerburgh, 2021). While Corbae and D'Erasmo (2021) build a structural model of banking industry dynamics, their focus is on regulatory policy. I instead explore the macroeconomic implications of bank market power through the lens of firm dynamics. Similarly, Elenev, Landvoigt, and Van Nieuwerburgh (2021) examine capital constraints among banks and firms, whereas I highlight a novel amplification channel: the endogenous dispersion of firm-specific loan markups. Another related contribution is Dempsey (2025), who develops a model with heterogeneous firms that can choose between intermediated and direct finance, while heterogeneous banks compete with one another and with the bond market. His framework features an endogenous comparative advantage of banks in lending relative to non-banks and emphasizes the role of capital requirements when bank and non-bank credit are imperfect substitutes. While Dempsey (2025) models a continuum of monopolistically competitive banks, my focus is on competition among non-atomistic banks, which gives rise to endogenous cross-sectional dispersion in loan markups and, in turn, shapes firm-level financing and investment decisions.

Other papers study bank market power in the context of monetary policy transmission (e.g., Wang, Whited, Wu, and Xiao, 2022; Jamilov and Monacelli, 2025), often relying on markups derived from constant elasticity of substitution (CES) frameworks. In contrast, I use a generalized Euler equation framework that allows for time-varying, endogenous firm-level markups driven by dynamic bank-firm interactions.

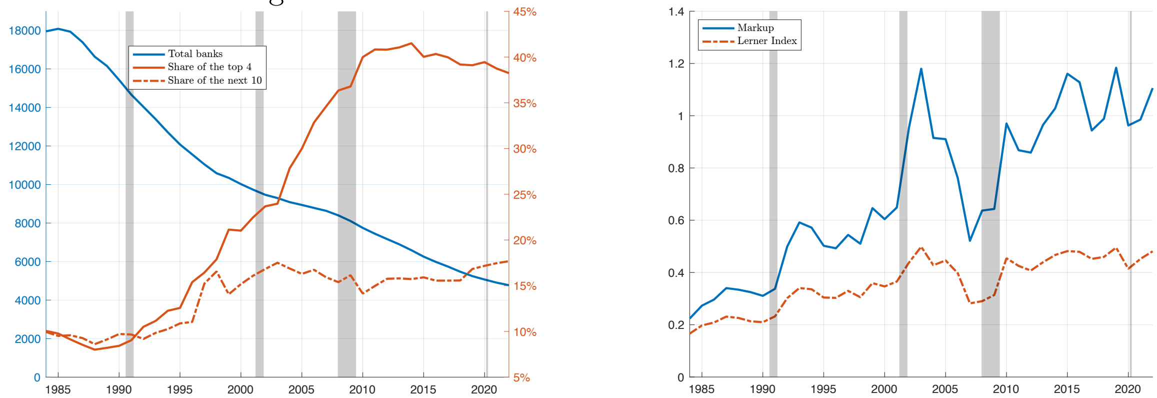
Outline. The rest of the paper is organized as follows. Section 2 reports three salient stylized facts. Section 3 presents a stylized version of the model aimed at delivering basic intuition about the proposed mechanism. Section 4 describes the infinite-horizon version of the model and discusses the main mechanism and its effects in detail. Section 5 explains the calibration and results of the oligopolistic stationary equilibrium. Section 6 illustrates the results of the three aforementioned macroeconomic shocks, such as the failure of one big bank and the entry of a new bank. Section 7 proposes an extension of the baseline model. Section 8 concludes.

2 Stylized Facts

The purpose of this section is to report the key empirical facts on bank market structure and market power. Appendix C.4 details the data and methodology and supplements the analysis with additional patterns.

(1) Bank Concentration and Markups Increased. I obtain bank-level balance-sheet and income-statement data for U.S. FDIC-insured commercial banks from the FDIC Banks API for the period 1984–2022, using year-end regulatory financial reports, and use these data to construct bank asset shares and markups in the spirit of De Loecker, Eeckhout, and Unger (2020) and Corbae and D'Erasmo (2021). Figure 1 shows that both the asset-market share of the top 4 banks and the asset-weighted markup have been increasing over time.

Figure 1. Bank Market Share and Markups



Notes: The left panel shows the evolution over time of the total number of FDIC-insured banks in the U.S., the asset-market share of the top 4 banks, and the asset-market share of the next 10 largest banks. The right panel shows the asset-weighted markup and Lerner index.

Bank-level markups are calculated as

$$\frac{\text{Interest Return on Loans}_t}{\text{Cost of Funds}_t + \text{Marginal Net Expenses}_t} - 1.$$

The interest rate on loans is calculated as the ratio between interest income on loans and the total loan amount. The cost of funds is calculated as the interest rate on both deposits and Fed funds, relative to the total of deposits and Fed funds. Marginal net expenses are defined as the marginal non-interest expenses net of marginal non-interest income. Marginal non-interest expenses and marginal non-interest income are derived from the trans-log functions (46) and (47) in Appendix C.4.2, along the lines of Demirgüç-Kunt and Martinez Peria (2010). The Lerner index is calculated as

$$\frac{\text{Markup}_t}{1 + \text{Markup}_t}.$$

Hence, a Lerner index of zero indicates the perfect competition benchmark. The Lerner index has been significantly above zero and increasing over time, pointing to an increase in market power. Appendix C.4.2 explains all these measures in detail.

(2) The Evolution of Bank Concentration is Positively Correlated with Credit Spreads. Using the same database as in point (1), I run the following bank-level regression:

$$\log(R_{b,L,t} - R_{M,t}) = \beta_0 + \beta_1 \times \log L_{b,t} + \beta_2 \times \text{Charge-off}_{b,t} + \beta_3 \times C_{5,t} + \alpha_b,$$

where α_b denotes bank fixed effects. The dependent variable is the log of the bank-level credit spread, $\log(R_{b,L,t} - R_{M,t})$, calculated as the difference between the loan rate, $R_{b,L,t}$, and the federal funds rate, $R_{M,t}$. The independent variables are (i) the log of the outstanding quantity of loans ($L_{b,t}$), (ii) the charge-off rates, and (iii) the concentration of the largest 5 banks ($C_{5,t}$). I present the results in Table C10 in Appendix C.4.2. The coefficient of interest is β_3 which is estimated to be .696 (statistically significant at the 1 percent level), hence implying that a one-percentage-point increase in $C_{5,t}$ results in a .696 percent increase in the annualized credit spread, ceteris paribus. Table C11 shows that the statistical significance of the coefficients is robust to clustering standard errors at the bank-year level.

(3) Controlling for Deal Amount and a Proxy for Corporate Default, Smaller Firms Tend to Pay Higher Credit Spreads. In order to investigate interest rates at a firm-bank-level, I use the Roberts Dealscan-Compustat Linking Database following Chava and Roberts (2008), which connects borrowers and lenders from 1986 to 2012. I run the following firm-bank-level regression:

$$\text{All-In Drawn}^d_{f,b,t} = \beta_0 + \beta_1 \times \log K_{f,t} + \beta_2 \times \log L^d_{f,b,t} + \beta_3 \times \text{Altman Z-score}_{f,t} + \gamma_t,$$

where All-In Drawn$^d_{f,b,t}$ indicates the interest rate spread on a facility d paid by a firm f to a bank b, $K_{f,t}$ indicates the book value of a firm f, $L^d_{f,b,t}$ indicates the facility amount, Altman Z-score is a proxy for firm f's bankruptcy probability, and γ_t indicates time fixed effects.[6] I present the results of this baseline specification and several robustness checks in Table C7 in Appendix C.4.1. The coefficient of interest is β_1 which is estimated to be $-.201$ (statistically significant at the 1 percent level), hence implying that a firm with a size corresponding to the 75$^{\text{th}}$ percentile pays about 60 basis points less in annualized credit spreads than a firm with a size corresponding to the 25$^{\text{th}}$ percentile, ceteris paribus. The estimate of β_1 remains negative and significant across specifications. Figure C13 reports the time fixed effect, γ_t, over time. On average, the estimated time fixed effects increase over time, which is consistent with the notion that rising bank concentration has exerted upward pressure on interest rates across all firms. Relatedly, I run the following additional regression which explicitly controls for the interaction between firm size and bank concentration:

$$\begin{aligned}\text{All-In Drawn}^d_{f,b,t} = \beta_0 + \beta_1 \times \log K_{f,t} + \beta_2 \times \log L^d_{f,b,t} + \beta_3 \times \text{Altman Z-score}_{f,t} + \beta_4 \times \text{C}_{5,t} + \\ \beta_5 \times (\log K_{f,t} \cdot \text{C}_{5,t}).\end{aligned}$$

The new coefficient of interest, β_5, is estimated to be negative, while β_4 is positive and β_1 remains negative. All coefficients are significant at the 1 percent level. This suggests that a higher level of concentration interacts with firm size by increasing interest rates disproportionately more for smaller firms. Table C7 in Appendix C.4.1 reports all estimated coefficients and their significance levels.

[6] Appendix C.4.1 explains the data and all these measures in detail.

3 Stylized Model

In this section, I analyze a two-period model designed to provide preliminary intuition for the infinite-horizon model presented in Section 4. An oligopolistic banking sector interacts with a continuum of heterogeneous firms in the presence of idiosyncratic productivity and default shocks. I provide analytical results on how increasing the number of banks B impacts various financial and macroeconomic variables, such as loans, interest rates, expected returns, investment, leverage, and TFP. In the stylized version of the model, there are two dates, denoted $t \in \{0, 1\}$. Note that, in the stylized model, the number of banks is exogenous, whereas in the infinite-horizon model it is endogenous (banks make entry and exit decisions).

Preferences. There is a continuum of identical households with preferences:

$$C_0 + \beta \cdot C_1, \tag{1}$$

where C_t is the household's consumption at time t and $\beta \in (0, 1)$ is the discount factor.

Technology. There is a continuum of firms $j \in [0, 1]$. In each period $t = 0, 1$, the output $y_t(j)$ produced by each firm j is given by the production function $y_t(j) = z_t(j) \cdot k_t(j)^\alpha$, where $0 < \alpha < 1$.

Ownership Structure. The households own all banks and firms. Each firm j is characterized by its state vector

$$x_t(j) \equiv \{\{l_{b,t}(j)\}_{b=1}^{B}, r_{l,t}(j), k_t(j), \mathcal{I}_t(j), z_t(j)\},$$

where $l_{b,t}(j)$ denotes the firm's loan from bank b, $r_{l,t}(j)$ (throughout the paper I will also refer to $R_{l,t}(j) = 1 + r_{l,t}(j)$) is the interest rate (charged by all banks), $k_t(j)$ is the firm's capital stock, $z_t(j)$ is the firm's productivity, and $\mathcal{I}_t(j)$ is an indicator function that takes the value 1 if the firm has not defaulted. Let $\phi(x_0)$ denote the density function of firms in the economy at $t = 0$.

Markets. There are five markets in the economy: bank debt, bank equity, firm loans, firm equity, and the market for the representative good.

Bank equity and debt markets. The household invests in the production sector by supplying equity or debt to banks and by supplying equity to the firms and faces budget constraints:

$$C_0 + \sum_{b=1}^{B} (p_b \cdot S_{b,1} + D_{b,1}) + \int p_0 \cdot S_1 \, \mathrm{d}\Phi = \sum_{b=1}^{B} (p_b + \pi_{b,0}) \cdot S_{b,0} + \int (p_0 + \tilde{d}_0) \cdot S_0 \, \mathrm{d}\Phi \tag{2}$$

$$C_1 = \sum_{b=1}^{B} (\pi_{b,1} \cdot S_{b,1} + R_D \cdot D_{b,1}) + \int \mathcal{I} \cdot \tilde{d}_1 \cdot S_1 \, \mathrm{d}\Phi, \tag{3}$$

where p_b, $S_{b,0}$, $S_{b,1}$, $D_{b,1}$, $\pi_{b,0}$, and $\pi_{b,1}$ are, respectively, the bank's share price, the share holdings at $t = 0, 1$, the debt holdings at $t = 1$, and the bank's profit at $t = 0, 1$. Bank b demands equity and debt from the household, in order to finance loans to firms.

Firm equity market. The household invests in the production sector by supplying equity to the firms, and faces budget constraints (2) and (3), where p_0, S_0, S_1, $\tilde{d}_0$, and $\tilde{d}_1$ are, respectively, the share price, share holdings at $t = 0, 1$, and the dividend of each firm (net of equity issuance cost) at $t = 0, 1$. Firms demand equity from, or distribute dividends to, the household. If a firm decides to issue equity, it incurs a quadratic equity issuance cost at $t = 0$ (with λ_0 being a positive constant):

$$\lambda(d_0) = \begin{cases} \lambda_0 \frac{d_0^2}{2} & \text{if } d_0 \leq 0 \\ 0 & \text{if } d_0 > 0 \end{cases}, \tag{4}$$

where d_0 is the firm's dividend at $t = 0$, defined below. The cost of equity issuance could stem from information frictions, such as signaling costs or adverse selection.

Firm loan market. A finite number B of (identical) banks supply loans to the continuum of firms. Each bank $b \in \{1, \dots, B\}$ can issue non-state-contingent loans $l_{b,1}$ to each firm. Loans are due for repayment in the next period, unless the firm defaults. A firm j takes the interest rate $r_1(j)$ as given and chooses how much to invest and how much to borrow from each bank. Banks take each firm's demand schedule as given and compete à la Cournot, i.e., simultaneously and independently choose their loan portfolios. The Cournot model has the convenient property that the number of banks is a sufficient statistic to determine market power, rendering it a parsimonious and tractable modeling choice. Alternatively, one could consider a Bertrand model where banks face capacity constraints, in order to rule out perfectly competitive outcomes. As shown by Kreps and Scheinkman (1983), quantity precommitment followed by Bertrand competition would induce Cournot outcomes.

Goods market. The representative household demands goods supplied by all firms.

Shocks. At time 0, firms are heterogeneous with respect to their capital stock k_0 and their idiosyncratic productivity z_0. At time 1, there are two types of idiosyncratic shocks: the firm can default, with exogenous probability $1 - \rho$ and, if it survives, z_1 realizes according to $\log z_1 = \rho_z \log z_0 + \xi_1$ where $\xi_1 \sim \mathcal{N}(0, \sigma_\xi^2)$ and $0 \leq \rho_z < 1$.

Timing. All decisions are made at $t = 0$. Given the initial distribution of firms with pdf $\phi(x_0)$ (and cdf $\Phi(x_0)$), the timing is as follows: (1) each firm produces output $y_0 = z_0 k_0^\alpha$; (2) each bank finances its supply of loans, $\int l_{b,1}(x_0)\,\mathrm{d}\Phi(x_0)$, by issuing equity and/or debt; (3) each firm takes the interest rate $R_1(x_0)$ as given and chooses how much to invest and the loan amount to demand from each bank; (4) banks take each firm's demand schedule as given and compete with each other to supply the loans. The outcome is a contract specifying the loan amount $l_{b,1}(x_0)$, the interest rate $R_1(x_0)$, and the new level of capital $k_1(x_0)$; and

(5) firms distribute dividends $d_0 = z_0 k_0^\alpha + (1-\delta)k_0 - k_1 + \sum_b^B l_{b,1}$ to the household.[7] At $t = 1$, a mass $1-\rho$ of defaulting firms exits the market. For the surviving firms, z_1 is realized and: (1) firms produce output $y_1 = z_1 k_1^\alpha$; (2) firms repay their outstanding debt plus interest $R_1(x_0) \cdot \sum_b^B l_{b,1}(x_0)$; (3) each bank distributes its profit $\int \rho R_1(x_0) l_{b,1}(x_0)\, \mathrm{d}\Phi(x_0)$ to the saver; and (4) firms distribute dividends $d_1 = z_1 k_1^\alpha + (1-\delta)k_1 - R_1 \sum_b^B l_{b,1}$ to the household.

3.1 Agents' Optimization Problems

The representative household maximizes its intertemporal utility (1) subject to the budget constraints (2) and (3), yielding Euler equations that pin down the price of bank equity $\forall b: \;\; \beta\pi_{b,1} = p_{b,0}$, and the price of each firm's equity

$$\rho\beta\mathbb{E}_0\left[\frac{d_1}{p_0}\right] = 1 - \lambda_d(d_0). \tag{5}$$

Firms maximize the net present value of dividends $d_0 + \beta \cdot \mathbb{E}_0\left[\mathcal{I} \cdot d_1\right]$. The firm's optimality condition with respect to the loan requires that the discounted expected future interest rate equal one net of the marginal equity issuance cost:

$$\rho\beta R_{l,1} = 1 - \lambda_d(d_0). \tag{6}$$

The firm's first-order condition with respect to capital requires that the future interest rate equals the expected marginal productivity of capital net of depreciation:

$$R_{l,1} = \mathbb{E}_0\left[1 + \alpha z_1 k_1^{\alpha-1} - \delta\right]. \tag{7}$$

Generalized Euler Equation. Bank strategies map firm characteristics (x_0) onto the current quantity and future interest rate of loans. Given the probability density function $\phi(x_0)$ and cumulative distribution function $\Phi(x_0)$, each bank b chooses $l_{b,1}(x_0)$ as a best response to the strategies of other banks, $l_{-b,1}(x_0)$, such that

$$\max_{l_{b,1}(x_0)} \quad \pi = -\int l_{b,1}(x_0)\, \mathrm{d}\Phi(x_0) + \beta \int \rho R_{l,1}(x_0) l_{b,1}(x_0)\, \mathrm{d}\Phi(x_0),$$

subject to equations (5)-(7) for all firms in the distribution.

Each bank's best response is characterized by the following generalized Euler equation (GEE)

$$\forall x_0: \;\; \frac{\partial \pi}{\partial l_{b,1}(x_0)} = -1 + \rho\beta\frac{\partial R_{l,1}(x_0)}{\partial l_{b,1}(x_0)} l_{b,1}(x_0) + \rho\beta R_{l,1}(x_0) = 0, \tag{8}$$

where $\frac{\partial R_{l,1}(x_0)}{\partial l_{b,1}(x_0)}$ can be determined by the implicit function theorem on equations (6) and (7).

[7] I assume for simplicity that firms start with $l_{b,0} = 0$.

Equation (8) is a generalized Euler equation because it contains the derivatives of each firm's policy functions. Each bank best responds by also internalizing how loans affect firm capital choice, $\frac{\partial k_1}{\partial l_{b,1}}$. For ease of notation, I drop the dependency of all optimal choices on x_0. The inverse elasticity $\frac{\partial R_{l,1}}{\partial l_{b,1}} \frac{l_{b,1}}{R_{l,1}}$ implicitly contained in the generalized Euler equation requires the determination of the term $\frac{\partial R_{l,1}}{\partial l_{b,1}}$. From equation (7)

$$\frac{\partial R_{l,1}}{\partial l_{b,1}} \frac{l_{b,1}}{R_{l,1}} = \mathbb{E}_0 \left[\underbrace{\alpha(\alpha-1) z_1 k_1^{\alpha-2}}_{\frac{\partial \mathrm{MPK}}{\partial k_1}} \cdot \frac{\partial k_1}{\partial l_{b,1}} \frac{l_{b,1}}{R_{l,1}} \right], \tag{9}$$

where $\mathrm{MPK} \equiv z_1 \cdot \alpha k_1^{\alpha-1}$. The formula suggests that bank market power depends on two components. First, banks extract higher markups from firms that exhibit a lower (i.e., more negative) derivative of the marginal productivity of capital $\frac{\partial \mathrm{MPK}}{\partial k_1}$.[8] These firms have less elastic credit demand because restrictions on loan supply reduce investment in physical capital and generate (i) a larger loss in production, since these firms have a high MPK; and (ii) a higher interest rate, since their capital levels correspond to a more concave region of the production function. Second, banks behave strategically by internalizing the effects of their actions on firm investment decisions, that is, on firm policy functions. This second effect is captured by the cross-elasticity $\frac{\partial k_1}{\partial l_{b,1}} \frac{l_{b,1}}{R_{l,1}}$. Finally, note that expressions for the two cross-derivatives can be found jointly by taking the total derivatives of equations (6) and (7):

$$\frac{\partial R_{l,1}}{\partial l_{b,1}} = \frac{1 - \rho\beta R_{l,1}}{\rho\beta l_{b,1}} \quad \text{and} \quad \frac{\partial k_1}{\partial l_{b,1}} = \frac{1 - \rho\beta R_{l,1}}{\rho\beta l_{b,1}} \cdot \frac{1}{\mathbb{E}_0 \left[\frac{\partial \mathrm{MPK}}{\partial k_1} \right]}.$$

In equilibrium, for the mass of financially constrained firms ($d_0 < 0$), the degree of imperfect competition (number of banks B) matters. For each firm, the equilibrium is a vector $(k_1^*, R_{l,1}^*, l_{b,1}^*, p_0^*)$ such that equations (5)-(8) hold. For the mass of firms that, in equilibrium, is not financially constrained, the degree of imperfect competition does not matter. For these firms, the solution is given by $(k_1^*, R_{l,1}^*, p_0^*)$ such that equations (5)–(7) hold, and the Modigliani-Miller theorem applies; hence, $l_{b,1}^*$ is undetermined. Note that, in the quantitative model, the household is risk-averse and a tax shield is present. These elements, combined with the non-atomistic behavior of banks, ensure that each firm's optimal capital structure is well defined in equilibrium.[9]

[8] Note that the term $\alpha - 1$ is always negative; hence, $\frac{\partial \mathrm{MPK}}{\partial k_1}$ and the inverse elasticity $\frac{\partial R_{l,1}}{\partial l_{b,1}} \cdot \frac{l_{b,1}}{R_{l,1}}$ are always negative. Banks exert higher market power when $\frac{\partial R_{l,1}}{\partial l_{b,1}} \cdot \frac{l_{b,1}}{R_{l,1}}$ is smaller.

[9] See Appendix B.2 for further details.

3.1.1 Characterization of the Equilibrium

I now describe intuitively the main mechanism that drives the analytical results presented in this section. First, equation (9) suggests that the higher the MPK of a firm, the higher the marginal value of one unit of loan for that firm that translates into a lower inverse elasticity $\frac{\partial R_{l,1}}{\partial l_{b,1}} \cdot \frac{l_{b,1}}{R_{l,1}}$. Second, the degree of imperfect competition (represented by the number of banks B) is relevant only for the mass of financially constrained firms. Note that being financially constrained is an equilibrium outcome, as it is determined by d_0, which is endogenous, and is itself influenced by the number of banks B. Consequently, banks endogenously exert a greater degree of market power on firms that, in equilibrium, become financially constrained and exhibit a high MPK. Intuitively, these firms have worse outside options (e.g., a high cost of non-bank finance) and one additional unit of investment in physical capital contributes significantly to their future production; hence, they exhibit a less elastic demand for credit. An imperfectly competitive financial sector internalizes that the same financial resources are more valuable for this type of firm and for their future growth paths; therefore, it can charge higher markups.

This mechanism leads financially constrained firms to grow more slowly in less competitive credit markets. As a result, the dispersion of marginal productivity of capital is higher when there are fewer banks B in the economy. At the same time, a lack of competition in financial intermediation reduces aggregate productivity, since firms grow toward their efficient level of capital on slower trajectories. This intuitive mechanism is the basis of Proposition I presented in Appendix B. In essence, a higher number of banks (i.e., a higher B) has the following effects: (i) aggregate loans per bank $\int l^*_{b,1}\, \mathrm{d}\Phi$ decreases; (ii) average loan interest rate $\int R^*_{l,1}\, \mathrm{d}\Phi$ decreases; (iii) aggregate physical investment $\int k^*_1 - (1-\delta)k_0\, \mathrm{d}\Phi$ increases; (iv) aggregate expected returns $\int \mathbb{E}_0 \left[d^*_1\right] / p^*_0\, \mathrm{d}\Phi$ decreases; (v) aggregate loans $\int \sum_b^B l^*_{b,1}\, \mathrm{d}\Phi$ increases; (vi) aggregate leverage $\int \sum_b^B l^*_{b,1}/k^*_1\, \mathrm{d}\Phi$ increases; (vii) aggregate TFP $\int k_1^{*\alpha}\, \mathrm{d}\Phi / \left(\int k^*_1\, \mathrm{d}\Phi\right)^{\alpha}$ increases; (viii) variance of capital $\int k_1^{*2}\, \mathrm{d}\Phi - (\int k^*_1\, \mathrm{d}\Phi)^2$ decreases; (ix) variance of loan interest rates $\int R^{*2}_{l,1}\, \mathrm{d}\Phi - (\int R^*_{l,1}\, \mathrm{d}\Phi)^2$ decreases; and (x) variance of expected returns $\int \left(\mathbb{E}_0 \left[d^*_1\right] / p^*_0\right)^2 \mathrm{d}\Phi - \left(\int \mathbb{E}_0 \left[d^*_1\right] / p^*_0\, \mathrm{d}\Phi\right)^2$ decreases.

4 Model

In the two-period model, bank choices are static. In the infinite-horizon model, each bank faces a dynamic problem that: (i) depends on its own future strategies and on the current and future strategies of other banks; and (ii) is subject to dynamic loan demand by firms. The current and future distributions of firms also matter. The equilibrium concept used in this section is a Markov perfect equilibrium (e.g., Maskin and Tirole, 2001). Specifically, I characterize the equilibrium using generalized Euler equations in a similar fashion to the optimal fiscal policy literature (see, for instance, Klein and Ríos-Rull, 2003; Klein, Krusell, and Ríos-Rull, 2008; and Clymo and Lanteri, 2020). In this section, I build a dynamic framework to study firm financing-investment decisions when banks are large, non-atomistic players, strategically interact with each other, and face idiosyncratic firm default risk.[10] Households derive utility from a non-durable consumption good, own the shares of the banks, and supply deposits. Banks issue debt and use both internal resources and debt financing to extend loans to firms. Firms make investment decisions, taking into account the fact that debt provides a tax shield and issuing new equity is increasingly costly. The key feature of the framework is the simultaneous presence of strategic interactions among financial institutions, general equilibrium, macroeconomic shocks, and heterogeneous firms. Note that each firm signs an idiosyncratic contract with the banks: in equilibrium, banks have different degrees of market power over each single firm as a function of its idiosyncratic characteristics.

I now describe the model and define the stationary oligopolistic equilibrium, in which all aggregate quantities and prices are constant over time. I overcome the computational challenge by proposing GEE-based algorithms to solve for the oligopolistic stationary equilibrium and related transitional dynamics in the presence of strategic interactions, general equilibrium, and heterogeneous firms. The algorithms are detailed in Appendix A.

4.1 Environment

Time is discrete $t = 0, 1, \ldots$ and the horizon is infinite. There is an endogenous number B of identical big banks in equilibrium and an endogenous number $\mathcal{B}$ of identical small banks in equilibrium. Unlike big banks, small banks operate under capacity constraints. All banks are owned by a continuum of identical and infinitely lived households, equivalent to one representative household. The household also owns a continuum of firms $j \in [0, 1]$.

Preferences. The household ranks streams of consumption C_t according to the following lifetime utility function:

$$\sum_{t=0}^{\infty} \beta^t \cdot u(C_t), \tag{10}$$

where $\beta \in (0, 1)$ is the household's discount factor, and $u_c > 0$, $u_{cc} < 0$.

Technology. In each period t, the output $y_t(j)$ produced by each firm $j \in [0, 1]$ is given by the production function $y_t(j) = Z_t \cdot z_t(j) \cdot k_t(j)^\alpha$, $0 < \alpha < 1$, where Z_t is an aggregate TFP shock (which is used in Section 6 in the form of an unexpected shock) and $z_t(j)$ indicates firm j's idiosyncratic productivity with characteristics specified below.

[10] Specifically, I interpret the financial intermediation sector as a succession of decision makers – one at each date t – without commitment to future realized quantities of loans supplied.

Ownership Structure. The household owns all the banks and the entire mass of firms $j \in [0,1]$. Each firm j is characterized by its state vector

$$x_t(j) \equiv \left\{ \{l_{b,t}(j)\}_{b=1}^{B_t}, \{l^s_{b,t}(j)\}_{b=1}^{\mathcal{B}_t}, r_{l,t}(j), k_t(j), \mathcal{I}_t(j), z_t(j) \right\},$$

where $l_{b,t}(j)$ denotes the firm's loan from bank $b = 1, ..., B$, $l^s_{b,t}(j)$ denotes the firm's loan from a small bank $b = 1, ..., \mathcal{B}$, $r_{l,t}(j)$ is the interest rate (charged by all banks), $k_t(j)$ is the firm's capital stock, $\mathcal{I}_t(j)$ is a dummy variable that takes the value one if the firm has not defaulted, and $z_t(j)$ is the firm's idiosyncratic productivity. Let $\phi_t(x_t)$ denote the density function of firms in each given period after firms default and reenter, with associated cumulative distribution $\Phi_t(x_t)$. The household can save through banks or through firm equity, in this latter case it incurs the equity issuance cost (4). Throughout the paper, I will omit the explicit dependency on j of the firm-specific variables contained in $x_t(j)$ for notational convenience.

Markets. There are six markets in the economy: bank debt, bank equity, firm loans, firm equity, the interbank market, and the market for the representative good.

Bank equity and debt markets. The household invests in the production sector by supplying equity or debt to banks and faces the budget constraint:[11]

$$\begin{aligned} C_t + \sum_{b=1}^{B_{t+1}} (p_{b,t} S_{b,t+1} + D_{b,t+1}) + \sum_{b=1}^{\mathcal{B}_{t+1}} \left(p^s_{b,t} S^s_{b,t+1} + D^s_{b,t+1} \right) = \int \mathcal{I}_t \tilde{d}_t + (1 - \mathcal{I}_t) \tilde{d}_0 \,\mathrm{d}j \\ + \sum_{b=1}^{\overline{B}_t} ((p_{b,t} + \pi_{b,t}) S_{b,t} + R_{D,t} D_{b,t}) + \sum_{b=1}^{\overline{\mathcal{B}}_t} \left((p^s_{b,t} + \pi^s_{b,t}) S^s_{b,t} + R^s_{D,t} D^s_{b,t} \right), \end{aligned} \tag{11}$$

where $p_{b,t}$, $S_{b,t}$, $S_{b,t+1}$, $D_{b,t}$, $D_{b,t+1}$, $R_{D,t}$, and $\pi_{b,t}$ are, respectively, the big bank's share price, the share holdings at t and $t+1$, the big bank's debt holdings at t and $t+1$, the interest rate on the bank's debt, and the big bank's profit at t. The corresponding variables denoted with suffix s refer to the small banks instead. Each bank b demands equity and debt from the representative household, in order to finance loans to firms.

Firm equity market. The household owns all the firms and, for firms that have not defaulted, it receives each firm's dividend $\tilde{d}_t$ (net of equity issuance cost) at t. Firms demand equity from, or distribute dividends to, the household. If a firm decides to issue equity, it incurs the equity issuance cost $\lambda(d_t)$ given by equation (4), where d_t is the firm's dividend at time t, defined below. The corresponding mass of firms that have defaulted is replaced by a mass of newly entering firms with characteristics x_0, which produce a dividend $\tilde{d}_0$. This dividend

[11] Note that $\overline{B}_t \equiv \max\{B_t, B_{t+1}\}$ and $\overline{\mathcal{B}}_t \equiv \max\{\mathcal{B}_t, \mathcal{B}_{t+1}\}$. If a bank decides to exit, then $p_{b,t} = 0$, but it still needs to be taken into account because it can liquidate its assets in the form of a dividend. If a bank decides to enter, it is owned by the household and requires an initial equity injection—i.e., a negative dividend—to operate.

can potentially be negative in the case of an equity issuance.

Firm loan market. The B_t big banks and the $\mathcal{B}_t$ small banks supply loans to a continuum of firms. Each bank $b \in \{1, \ldots, B_t\}$ and $b \in \{1 \ldots \mathcal{B}_t\}$ can issue non-state-contingent loans $l_{b,t+1}$ and $l^s_{b,t+1}$ to each firm. Loans are due for repayment in the next period, unless the firm defaults. A firm j takes the interest rate $r_{l,t+1}(j)$ as given and chooses how much to invest and how much to borrow from each bank. Banks take each firm's demand schedule as given and compete dynamically à la Cournot. Unlike big banks, small banks are subject to capacity constraints. The process determines the total amount of loans banks supply to each firm which, together with the firm's demand schedule, pins down the firm-specific interest rate $r_{l,t+1}(j)$. At time t, each bank and firm commit to such an interest rate.

Interbank market. A big bank b can lend $M_{b,t}$ to other big banks that will be repaid in the following period at rate $r_{M,t+1}$. Since all big banks are identical, in equilibrium $\forall b: \ M_{b,t} = 0$. Similarly, a small bank b can lend $M^s_{b,t}$ to other small banks that will be repaid in the following period at rate $r_{M,t+1}$. Since all small banks are identical, in equilibrium $\forall b: \ M^s_{b,t} = 0$. For simplicity, I assume lending between big and small banks is not allowed.

Goods market. The representative household demands goods supplied by all firms.

Shocks. At time t, $\mathcal{I}_t$ takes the value of one with probability ρ_t, and zero with probability $1-\rho_t$. A new mass of firms reenters the economy so that the total mass remains constant over time. I relax this assumption in Subsection 6.2. Reentering firms are assigned characteristics x_0, and I set their initial capital k_0 as described in Section 5. I assume that, in the event of default, the economy recovers $\nu k + k_0$ from the defaulting firm's capital. Banks receive the portion νk, allocated across banks in proportion to their outstanding loan shares. The portion k_0 is transferred to the household, the owner of both firms and banks, and is recycled to finance newly entering firms. This assumption ensures that new firm entry is not fueled by exogenous resource injections, but rather by reallocating existing capital within the economy. For simplicity, I set the initial amount of the loan owed to the big bank as $l_{b,0} = 0$ for all $b \in \{1, \ldots, B\}$, and the initial amount owed to the small bank as $l^s_{b,0} = 0$ for all $b \in \{1, \ldots, \mathcal{B}\}$. Firms have heterogeneous productivity $z_t(j)$, with $z_0(j)$ drawn at the birth of the firm from a Gaussian probability distribution function with mean 1 and variance σ_z^2.[12] Finally, in Section 7, I also propose and solve an extension of the model with firm endogenous default decisions, which yields results consistent with the baseline model.

Government. The government imposes proportional taxes τ on firm production. Firms can deduct loan interest and depreciated capital from their taxes. The government runs a

[12] Similarly in spirit to Melitz (2003) and Asplund and Nocke (2006).

balanced budget constraint. That is, the government uses the aggregate revenue from taxes

$$T_t = \tau \left[\int \mathcal{I}_t \left(Z_t z_t k_t^{\alpha} - \sum_{b=1}^{B_t} r_{l,t} l_{b,t} - \sum_{b=1}^{\mathcal{B}_t} r_{l,t} l_{b,t}^{s} - \delta k_t \right) \mathrm{d}j + \int (1 - \mathcal{I}_t) \left(Z_t z_0 k_0^{\alpha} - \delta k_0 \right) \mathrm{d}j \right],$$

to finance an exogenous government expenditure that exactly balances T_t at each point in time.

Timing. The aggregate state of the economy at time t is

$$X_t \equiv \{\{D_{b,t}\}_{b=1}^{B_t}, \{D_{b,t}^{s}\}_{b=1}^{\mathcal{B}_t}, r_{D,t}, \{M_{b,t}\}_{b=1}^{B_t}, \{M_{b,t}^{s}\}_{b=1}^{\mathcal{B}_t}, r_{M,t}, B_t, \mathcal{B}_t, Z_t, \phi_t(x_t)\}.$$

Given X_t, the timing is as follows: (1) a mass $1-\rho_t$ of firms defaults, (2) each surviving firm produces output $y_t = Z_t z_t k_t^{\alpha}$ and repays its debt $\sum_{b}^{B_t} R_{l,t} l_{b,t} + \sum_{b}^{\mathcal{B}_t} R_{l,t} l_{b,t}^{s}$ to all incumbent banks; (3) a mass $1-\rho_t$ of firms enter the production sector with characteristics x_0; (4) each incumbent bank decides whether to exit; (5) potential bank entrants decide whether to enter, this requires an initial equity injection from the household; (6) each bank finances its supply of loans, $\int l_{b,t+1}(x_t)\,\mathrm{d}j$, by issuing equity and/or debt; (7) each firm takes the interest rate $r_{l,t+1}(x_t)$ as given and chooses how much to invest and the loan amount to demand from each bank; (8) banks take each firm's demand schedule as given and compete with each other to supply the loans. The outcome is a contract specifying the loan amount $l_{b,t+1}(x_t)$, the interest rate $r_{l,t+1}(x_t)$, and the new level of capital $k_{t+1}(x_t)$; (9) firms distribute dividends $d_t = (1-\tau)\left[Z_t z_t k_t^{\alpha} - \sum_{b}^{B_t} r_{l,t} l_{b,t} - \sum_{b}^{\mathcal{B}_t} r_{l,t} l_{b,t}^{s} \right] + \tau \delta k_t - \tilde{i}_t$ to the household, where $\tilde{i}_t = i_t + \sum_{b=1}^{B_t} l_{b,t} - \sum_{b=1}^{B_{t+1}} l_{b,t+1} + \sum_{b=1}^{\mathcal{B}_t} l_{b,t}^{s} - \sum_{b=1}^{\mathcal{B}_{t+1}} l_{b,t+1}^{s}$ and i denotes investment in physical capital; (10) bank b distributes profit to the household. To simplify notation, I omit explicit time subscripts from variables throughout the remainder of the model exposition. It is understood that (x, X) refers to (x_t, X_t), and (x', X') refers to (x_{t+1}, X_{t+1}).

4.2 Household

I now describe the household's problem in recursive form. Let $V_H(X)$ be the value function of the household with bank debt holdings $\mathbf{D} = [D_1 \ldots D_B]$, and bank equity holdings $\mathbf{S} = [S_1 \ldots S_B]$. This function satisfies the following functional equation:

$$V_H(X) = \max_{\mathbf{S}', \mathbf{D}'} \quad u(C) + \beta \cdot V_H(X') \tag{12}$$

subject to the budget constraint (11). The left-hand side of the budget equation (11) reports the household's expenditures: aggregate consumption and purchases of equity and debt issued by banks. The right-hand side reports the household's resources: returns on previously held equity and debt issued by each bank b, together with firm dividends.

The household takes the future bank debt rate r'_D and future bank profits as given, and chooses bank debt and bank equity holdings according to:

$$\forall b: \quad 1 = \mathcal{M}'(X, X') \cdot \frac{p'_b + \pi'_b}{p_b} \tag{13}$$

$$\forall b: \quad 1 = \mathcal{M}'(X, X') \cdot R'_D, \tag{14}$$

where $\mathcal{M}' \equiv \beta \frac{u_c(C')}{u_c(C)}$, and π_b is the profit of bank b distributed as a dividend to the household.

4.3 Firms

I now characterize firm j's problem in recursive form. For convenience, I omit the index notation j, i.e., all idiosyncratic time t state variables $x(j) \equiv \{\{l_b(j)\}_{b=1}^{B}, \{l_b^s(j)\}_{b=1}^{\mathcal{B}}, r_l(j), k(j), z(j)\}$ and corresponding choices are denoted without index j. Let $V_F(x, X)$ be the value function of the firm j with loan holdings $[l_1 \dots l_B]$ and $[l_1^s \dots l_{\mathcal{B}}^s]$ from each bank and capital k. This function satisfies the following functional equation:

$$V_F(x, X) \;=\; \max_{\{l'_b\}_{b=1}^{B'}, \{l_b^{s\prime}\}_{b=1}^{\mathcal{B}'}, k'} \; d - \lambda(d) + \mathbb{E}\left[\mathcal{I}' \cdot \mathcal{M}'(X, X') \cdot V_F(x', X') \mid (x, X)\right],$$

subject to

$$\begin{aligned}
k' &= k(1-\delta) + i \\
i &= \tilde{i} - \sum_{b=1}^{B} l_b + \sum_{b=1}^{B'} l'_b - \sum_{b=1}^{\mathcal{B}} l_b^s + \sum_{b=1}^{\mathcal{B}'} l_b^{s\prime} \\
d &= (1-\tau)\left[Zzk^{\alpha} - \sum_{b=1}^{B} r_l l_b - \sum_{b=1}^{\mathcal{B}} r_l l_b^s\right] + \tau\delta k - \tilde{i} \\
\lambda(d) &= \mathbb{I}[d \le 0] \cdot \lambda_0 \frac{d^2}{2} + \mathbb{I}[d > 0] \cdot 0,
\end{aligned}$$

where $\mathcal{M}'$ is the discount factor of the household, as described in the previous subsection, and $\mathbb{I}$ is an indicator function that takes the value one when the argument is true and zero otherwise. Each firm takes the future loan-market rate r'_l as given and finances itself through internal financing (production and equity issuance) and external financing (loans from banks). The first-order condition with respect to k' is

$$1 - \lambda_d(d) = \mathbb{E}\left[\mathcal{I}' \cdot \mathcal{M}'(X, X') \cdot \left(1 + (1-\tau)\left(Z'z'\alpha k'^{\alpha-1} - \delta\right)\right) \cdot (1 - \lambda_d(d')) \mid (x, X)\right], \tag{15}$$

which weighs the marginal cost of investing in one unit of physical capital today against the marginal benefit of the expected discounted future marginal production net of depreciation, both adjusted by the marginal cost of equity issuance in case dividends are negative. The first-order condition with respect to $l_b^{\chi\prime}$, with $\chi = \{., s\}$, is

$$1 - \lambda_d(d) = \mathbb{E}\left[\mathcal{I}' \cdot \mathcal{M}'(X, X') \cdot (1 + (1-\tau)r_l') \cdot (1 - \lambda_d(d')) \mid (x, X)\right], \tag{16}$$

which weighs the marginal benefit of borrowing one unit of loans today against the expected discounted marginal cost of the future interest repayment net of the tax shield, both adjusted by the marginal cost of equity issuance in case dividends are negative.

4.4 Banks

First, I present the decision problem of the incumbent banks. Then, I present the decision problem of the new potential entrants. Banks can be either big or small, i.e. $\chi = \{., s\}$. Consistently with the notation adopted above, big banks do not have a suffix whereas small banks have a suffix s. Note that banks are subject to aggregate constraints that contain the entire distribution of firms. In equilibrium, and given the exit shocks, there is a mapping between a firm's age age(j) and the vector $\{\{l_b(j)\}_{b=1}^{B}, \{l_b^s(j)\}_{b=1}^{\mathcal{B}}, r_l(j), k(j), \mathcal{I}(j)\}$. Hence, for computational purposes, it is convenient to represent the state space simply as $x(j) \equiv \{\text{age}(j), z(j)\}$. Given the exit shocks, the distribution of firms is simply given by $\phi(\text{age}, z)$, which is not directly affected by the actions of the banks.

4.4.1 Incumbents

An incumbent big bank $b = [1...B]$, or an incumbent small bank $b = [1...\mathcal{B}]$, chooses the new level of debt to demand from the household and the new level of loans to offer to each firm. Formally, the strategy space is defined as:

$$\mathcal{S}_b^{\chi\prime}(x, X) \equiv \{D_b^{\chi\prime}(X), l_b^{\chi\prime}(x, X)\}.$$

The new amount of debt issued ($\Delta D_b^{\chi\prime} = D_b^{\chi\prime} - D_b^{\chi}$) and internal financing F^{χ} is chosen to provide enough coverage for the change in interbank lending and aggregate loans:

$$F^{\chi} + \Delta D_b^{\chi\prime} = \Delta M_b^{\chi\prime} - \int \mathcal{I} \cdot l_b^{\chi} + (1-\mathcal{I}) \cdot \mathcal{R}^{\chi} - \left(\mathcal{I} \cdot l_b^{\chi\prime}(x, X) + (1-\mathcal{I}) \cdot l_b^{\chi\prime}(x_0, X)\right) \mathrm{d}j, \tag{17}$$

where I assume that, in the event of a firm's default, each bank recovers a fraction of the outstanding debt proportional to the firm's capital and to its share of total debt holdings. Specifically, $\mathcal{R}^{\chi} = \nu k \cdot l_b^{\chi} / (\sum_{b=1}^{B} l_b + \sum_{b=1}^{\mathcal{B}} l_b^s)$, so that the total recovery accruing to banks equals νk. The additional portion k_0 is transferred to the household, the owner of both firms and banks, and recycled to finance newly entering firms. This assumption ensures that new firm entry is not fueled by exogenous resource injections, but rather by reallocating existing capital within the economy. Since banks have no control over this amount, it does not appear in their dividend payouts or generalized Euler equations. I now describe the bank's problem in recursive form. Let $V_b^{\chi}(X)$ be the value function of a bank b. This function satisfies the following functional equation:

$$\tilde{V}_b^{\chi}(X) = \max_{\{D_b', r_D'\},\ M_b',\ \{l_b^{\chi\prime}(x,X), r_l'(x,X)\}} \quad \pi_b^{\chi} + \mathcal{M}'(X, X') \cdot V_b^{\chi}(X') \tag{18}$$

where each incumbent bank makes an exit decision

$$V_b^{\chi}(X) = \max\{0, \tilde{V}_b^{\chi}(X)\} \tag{19}$$

subject to: (i) equation (17), (ii) the household's interest rate-quantity schedule jointly defined by equations (13) and (14), and (iii) each firm's interest rate-quantity schedule jointly defined by equations (15) and (16). Bank b's profit π_b is given by:

$$\pi_b^{\chi} = \int \mathcal{I} \cdot r_l \cdot l_b^{\chi}\, \mathrm{d}j + r_M M_b^{\chi} - r_D D_b^{\chi} - F^{\chi}. \tag{20}$$

Future market rates $r_D'(X)$ and $r_l'(x, X)$ adjust consistently with the interest rate-quantity schedules. Each bank b issues bank debt according to a generalized Euler equation:

$$1 = \mathcal{M}'(X, X') \cdot R_D'(X, X') \cdot \left(1 + \eta_D^{\chi\prime}(X, X')\right), \tag{21}$$

where $\eta_D^{\chi\prime}$ is the inverse elasticity $\frac{\partial R_D'}{\partial D_b^{\chi\prime}} \cdot \frac{D_b^{\chi\prime}}{R_D'}$ between debt and its rate. In principle, equation (21) is a best response function that captures the trade-off that a bank faces issuing new debt. Every new unit of debt increases the current investing capacity but needs to be repaid tomorrow at the contracted interest rate. Moreover, since $\eta_D^{\chi\prime}$ is non-negative, when a bank issues new debt it also increases the market rate of deposits, incurring an additional future marginal cost. In equilibrium, $\eta_D^{\chi\prime}$ is zero, as implied by equations (21) and (14). Without aggregate risk, households are indifferent to the financing structure of the banks. Small banks are identical to big banks except that their loan-supply choice is subject to a capacity constraint. Specifically, for a small bank $b \in [1...\mathcal{B}]$ and a big bank $\tilde{b} \in [1...B]$, the capacity constraint reads:

$$\forall (b, \tilde{b}) :\ l_b^{s\prime} \leq \kappa \cdot l_{\tilde{b}}', \tag{22}$$

which, intuitively, captures the idea that small banks face steep scaling costs, represented by a fraction $0 < \kappa < 1$ relative to the size of the large banks. Note that, in equilibrium, $\forall (b, \tilde{b}) :\ l_b^{s\prime} = \kappa \cdot l_{\tilde{b}}'$, since small banks are identical to big banks except the capacity constraint. Hence, if given the opportunity, small banks would grow to the same size as big banks. In equilibrium, all firms borrow from all banks. Although a firm may borrow different amounts from large and small banks, the borrowing proportions are determined solely by the capacity constraint of small banks. On the bank side, there is a generalized Euler equation (23) for each firm type in the distribution with characteristics x, which closes the model and, when taken together with the firm's optimization problem, determines the firm-specific big-bank loan level, interest rate, and new level of physical capital. Given this, the allocation between big and small banks is determined by the small bank capacity constraint (22). An interesting avenue for future research is to allow firms to choose from which banks to borrow based not only on capacity constraints, but also on additional considerations—such as the probability of default of each bank.[13]

The generalized Euler equation that arises from the big bank loan first-order condition is

$$1 = \mathbb{E}\left[\mathcal{M}'(X,X') \cdot \left(\mathcal{I}' \cdot R_l'(x,X,x',X') \cdot (1+\eta_l'(x,X,x',X')) + (1-\mathcal{I}') \cdot \frac{\partial \mathcal{R}'}{\partial {l_b}'}\right) \mid (x,X)\right], \tag{23}$$

and, similarly, the first-order condition for small banks is

$$1 + \mathcal{Z} = \mathbb{E}\left[\mathcal{M}'(X,X') \cdot \left(\mathcal{I}' \cdot R_l'(x,X,x',X') \cdot \left(1+\eta_l^{s\prime}(x,X,x',X')\right) + (1-\mathcal{I}') \cdot \frac{\partial \mathcal{R}^{s\prime}}{\partial l_b^{s\prime}}\right) \mid (x,X)\right], \tag{24}$$

where $\mathcal{Z}$ is the Lagrange multiplier on (22) and $\eta_l^{\chi\prime}(x,X,x',X') \equiv \frac{\partial R_l'}{\partial l_b^{\chi\prime}} \cdot \frac{l_b^{\chi\prime}(x,X)}{R_l'(x,X,x',X')}$ is the firm-specific inverse elasticity between loans and their rates. Equations (23) and (24) are functional equations that depend on the idiosyncratic characteristics of each firm. In equilibrium, equation (24) is unnecessary since equation (22) is binding, as discussed above. Nevertheless, the presence of small banks is important in shaping the inverse elasticity since it enters directly in $\eta_l^{\chi\prime}(x,X,x',X')$ and $\mathcal{M}'(X,X')$. See Subsection 4.4.2 for details on how to calculate $\eta_l^{\chi\prime}(x,X,x',X')$.

Equations (23) and (24) are the best response functions that capture the trade-off that a bank faces issuing a new unit of loan to a specific firm. Every new unit of loan decreases the current bank's dividend but produces a marginal income tomorrow at the contracted interest rate. Moreover, since $\eta_l^{\chi\prime}(x,X,x',X')$ is non-positive, when banks issue new loans they are also decreasing the future market rate of loans, incurring a marginal loss in the future. Note that banks best respond by internalizing the effects that their actions have on aggregate quantities and all firm choices (e.g., if a bank changes the quantity of loan offered to a firm, that firm might decide to re-optimize and adopt a different capital structure as a function of the credit market conditions). All these equations, together with an Euler equation that

[13] For instance, if certain firms borrow disproportionately more from smaller banks (which may be more likely to fail), this could create interesting spillover effects.

governs bank behavior in the interbank market

$$1 = \mathcal{M}'(X, X') \cdot R'_M(X, X'), \tag{25}$$

capture the decision-making behavior of each bank. The outcome of the game played by the banks at time t is a contract that pins down the firm-specific intermediation margin $R'_l(x, X, x', X') - R'_D(X, X')$. In principle, this margin can be decomposed into: (i) firm-specific loan intermediation margin $(R'_l(x, X, x', X') - R'_M(X, X'))$ and (ii) debt intermediation margin $(R'_M(X, X') - R'_D(X, X'))$. Note that the debt intermediation margin is zero, since $\eta_D^{\chi\,\prime} = 0$, hence $R'_D(X, X') = R'_M(X, X')$.

4.4.2 Calculation of the Inverse Elasticity η_l^χ

This subsection contains the calculation of the inverse elasticity η_l^χ in the generalized Euler equations (23) and (24). First, combine the firm first-order conditions (15) and (16) to define the two functions:

$$\begin{aligned}
f\,(x, X, x', X') &\equiv Z'z'\alpha k'^{\alpha-1} - \delta - R'_l + 1 = 0, \\
g\,(x, X, x', X') &\equiv \rho \cdot \mathcal{M}' \cdot (1 - \lambda_d^{\chi'})\,((1-\tau)(R'_l - 1) + 1) - 1 + \lambda_d^{\chi} = 0.
\end{aligned}$$

Second, compute the total derivatives of these two functions with respect to k' and $l_b^{\chi'}$. Hence, solve the resulting linear system to get the following expression:

$$\frac{\partial R'_l}{\partial l_b^{\chi'}} = \frac{f_{k'} \cdot g_{l_b^{\chi'}} - f_{l_b^{\chi'}} \cdot g_{k'}}{f_{R'_L} \cdot g_{k'} - f_{k'} \cdot g_{R'_L}}.$$

Note that $f_{l_b^{\chi'}} = 0$, $f_{R'_l} = -1$, and $f_{k'} = Z'z'\alpha(\alpha - 1)k'^{\alpha-2}$. Therefore, the elasticity $\eta_l^{\chi'}$ is given by

$$\eta_l^{\chi'} = \frac{\partial R'_l}{\partial l_b^{\chi'}} \frac{l_b^{\chi'}}{R'_l} = -\frac{f_{k'} \cdot g_{l_b^{\chi'}}}{g_{k'} + f_{k'} \cdot g_{R'_l}} \frac{l_b^{\chi'}}{R'_l}, \tag{26}$$

where the partial derivatives of the g function are

$$g_{k'} = \rho \frac{\partial \mathcal{M}'}{\partial k'} (1 - \lambda_d(d'))\,((1-\tau)r_l{}' + 1) - \rho \mathcal{M}' \frac{\partial \lambda_d(d')}{\partial k'}\,((1-\tau)r'_l + 1) + \frac{\partial \lambda_d(d)}{\partial k'}, \tag{27}$$

$$g_{l_b^{\chi'}} = \rho \frac{\partial \mathcal{M}'}{\partial l_b^{\chi'}} (1 - \lambda_d(d'))\,((1-\tau)r_l{}' + 1) - \rho \mathcal{M}' \frac{\partial \lambda_d(d')}{\partial l_b^{\chi'}}\,((1-\tau)r_l{}' + 1) + \frac{\partial \lambda_d(d)}{\partial l_b^{\chi'}}, \tag{28}$$

$$g_{R'_l} = \rho \frac{\partial \mathcal{M}'}{\partial R'_l} (1 - \lambda_d(d'))\,((1-\tau)r_l{}' + 1) - \rho \mathcal{M}' \frac{\partial \lambda_d(d')}{\partial R'_l}\,((1-\tau)r_l{}' + 1) + \rho \mathcal{M}'\,(1-\tau) + \frac{\partial \lambda_d(d)}{\partial R'_l}. \tag{29}$$

Equations (27)-(29) contain the strategies of all other banks $\mathbf{l}^{\chi}_{-b} = [l^{\chi}_1, ..., l^{\chi}_B]\backslash\{l^{\chi}_b\}$ in the discount factor $\mathcal{M}'$ and its derivatives, in the marginal equity issuance costs $\lambda_d(d)$ and $\lambda_d(d')$, and also the derivatives of the future firm policy functions inside the derivatives of $\mathcal{M}'$ and inside the derivatives of $\lambda_d(d')$. Since all bank strategies enter $\mathcal{M}'$ via aggregate consumption, $\mathcal{M}'$ serves as a key channel for strategic interactions. In the stationary equilibrium, $\mathcal{M}' = \beta$; however, for any $\gamma > 0$, its derivatives remain nonzero (see Appendix B.2 for details). This underscores the essential role of utility curvature: without it, aggregate consumption would not enter the stochastic discount factor and these strategic channels would vanish.

4.4.3 Entry Conditions

The new numbers of banks, B' and $\mathcal{B}'$, are the largest integers such that

$$\mathcal{M}(X, X') \cdot V^{\chi}_b(X') \geq F^{\chi}_E \quad \text{for } \chi \in \{., s\}, \tag{30}$$

where F^{χ}_E denotes the fixed entry cost for big ($\chi = .$) and small ($\chi = s$) banks, respectively. The updated numbers of banks satisfy $B' = B - B^{\text{Default}} + B^{\text{Entry}}$ and $\mathcal{B}' = \mathcal{B} - \mathcal{B}^{\text{Default}} + \mathcal{B}^{\text{Entry}}$, and are components of the vector X'. Here, $(B^{\text{Default}}, \mathcal{B}^{\text{Default}})$ are the numbers of big and small banks that default, as determined by conditions (19), and $(B^{\text{Entry}}, \mathcal{B}^{\text{Entry}})$ are the numbers of new entrants. Entry of a new bank requires an initial capital injection from the household.

4.5 Oligopolistic Equilibrium

The aggregate resource constraint of the economy is:

$$C + \int (I + \Lambda)\, \mathrm{d}j + T = \int Y\, \mathrm{d}j + \int R^{\text{rec}}\, \mathrm{d}j, \tag{31}$$

where $I \equiv \mathcal{I} \cdot i(x, X) + (1 - \mathcal{I}) \cdot i(x_0, X)$, $\Lambda \equiv \mathcal{I} \cdot \lambda(x, X) + (1 - \mathcal{I}) \cdot \lambda(x_0, X)$, $Y \equiv \mathcal{I} \cdot Zzk^{\alpha} + (1 - \mathcal{I}) \cdot Zz_0k_0^{\alpha}$, and $R^{\text{rec}} \equiv (1 - \mathcal{I}) \cdot \left(\sum_{b=1}^{B} \mathcal{R} + \sum_{b=1}^{\mathcal{B}} \mathcal{R}^s + k_0\right)$. The total production of the economy on the right-hand side of equation (31) can: (i) be consumed by the household, (ii) be used for aggregate investment in physical capital (in case some dividends are negative, some resources are spent to pay the equity issuance cost λ), and (iii) be paid in taxes. Note also that the right-hand side includes the resources recovered from the mass of defaulting firms, captured by the term R^{rec}.

A formal definition of the notion of *Recursive Stationary Oligopolistic Equilibrium* is presented in Definition 4.1, and its extension to the dynamic case is discussed in Section 6.

Definition 4.1. A **Recursive Stationary Oligopolistic Equilibrium** is a Markov perfect equilibrium where i) bank debt positions $(\{D_b\}^B_{b=1}, \{D^s_b\}^{\mathcal{B}}_{b=1})$ and the corresponding market rate R_D; ii) household holdings of bank shares, $(\{S_b\}^B_{b=1}, \{S^s_b\}^{\mathcal{B}}_{b=1})$ and the relative market prices $(\{p_b\}_{b=1}^{B}, \{p_b^s\}_{b=1}^{\mathcal{B}})$; iii) the interbank debt holdings $(\{M_b\}_{b=1}^{B}, \{M_b^s\}_{b=1}^{\mathcal{B}})$ and the relative market rate R_M; iv) the household's consumption C; v) the distribution $\phi(x)$; vi) the policy functions: $k'(x)$, $l'(x)$, $l^{s\prime}(x)$, and $R_l'(x)$; vii) the number of banks, $(B, \mathcal{B})$, is such that i) the household's problem is solved–i.e., equations (13) and (14) hold; ii) each firm's problem is solved–i.e., equations (15) and (16) hold; iii) each incumbent bank best responds to the strategies of all other banks–i.e., equations (21), (23), (24), and (25) hold; iv) it is not profitable for new banks to enter the intermediation market as per entry conditions (30) and incumbent banks have no incentive to default; v) all markets clear: 1) the good market clears –i.e., equation (31) holds; 2) each bank's equity market clears –i.e., $\forall b : S_b^{\chi} = 1$; 3) the interbank market clears –i.e., $\forall b : M_b^{\chi} = 0$.

5 Calibration

The choices of parameter values are described below and summarized in Table C1.

Household Preferences. A period in the model coincides with a quarter, consistent with the frequency in the data. I set $\beta = 0.995$ to match a quarterly stationary bank debt rate (or interbank debt rate) of 0.50%, which is consistent with the average of the 3-month T-bill rates calculated between 1997 and 2017. I use a constant relative risk aversion utility function with degree of relative risk aversion $\gamma = 1$. Hence, $u(C) = \log(C)$.

Firms. The quarterly depreciation rate δ is set to a standard value of 3%. In the baseline, I set the curvature of the reduced-form profit function α to 0.34.[14] I then conduct a sensitivity analysis by increasing θ to 0.75, which yields $\alpha = 0.51$, holding all other parameters constant. In this case, the banking sector profitability rises to 26.4%, compared to 16.5% under the baseline value of $\alpha = 0.34$. Bank profitability increases further to 30.5% when I raise θ to approximately 0.777, which implies $\alpha = 0.55$.[15] As shown in Figure C7 in Appendix C.1, when $\alpha = 0.51$, banks maintain positive markups across a broader segment of the firm distribution—particularly among lower-productivity firms. As α increases, the marginal return on capital is lower for very small firms but diminishes at a slower rate as capital rises. This implies that banks internalize a shift in marginal value: one dollar lent to a tiny firm yields less marginal return when α is higher, but becomes more valuable as the firm accumulates capital. This analysis suggests that the baseline value $\alpha = 0.34$ is conservative in terms of implied bank markups. See Appendix B.5 for further details.

[14] This value can be obtained using the mapping implied by a standard Cobb–Douglas production function with decreasing returns to scale. Specifically, consider the functional form $Zz^{v}(k^{a}n^{1-a})^{\theta}$, where k is capital, n denotes labor input (not explicitly modeled in this paper), $a = 0.35$ is the capital share, and $\theta = 0.6$ governs returns to scale. Defining $v = 1 - (1-a)\theta$, the reduced form curvature of profits with respect to capital is $\alpha = \frac{a\theta}{v} = \frac{0.35\times0.6}{1-0.65\times0.6} \approx 0.34$.

[15] These values of α are broadly consistent with estimates reported in the literature—for example, Cooper and Haltiwanger (2006) and Hennessy and Whited (2007) document values of 0.592 and 0.627, respectively.

The firm income tax rate is set to 19.7%, which was the effective corporate income tax in the U.S. in 2021. This tax rate is also broadly consistent with the ratio between taxes on corporate income over corporate profit (both time series are available from FRED) between 1997 and 2017. I use the net charge-off rates on Commercial and Industrial Loans (C&I) to pin down the risk of default, so that the stationary equilibrium value of ρ is $\bar{\rho} = 1 - 0.21\%$ (quarterly) consistent with the average of the time series between 1997 and 2017.[16] The equity issuance cost λ_0 is set to 0.8 to match an annualized frequency of equity issuance of 0.042. The frequency of equity issuance is computed from a sample of non-financial, unregulated firms from Compustat. Several studies (e.g., Gomes, 2001 and Hennessy and Whited, 2007) incorporate an equity issuance cost that includes both fixed and proportional components. In Appendix C.2, I explore an alternative functional form for the cost of equity issuance, which incorporates a fixed cost through a logit function to ensure differentiability. In each period, I also assume that a new mass of firms replaces the mass of firms that defaulted $1 - \bar{\rho}$, starting from age zero with capital k_0, which is set jointly with other parameters as explained below. In Subsection 6.2, I consider an extension where firm entry changes over time and does not necessarily match the firm exit mass.

Note that the parameter k_0, which governs the initial capital stock of reentering firms, plays an important role in both the recovery mechanism and the determination of firm leverage. In the event of default, the economy recovers $\nu k + k_0$ from the defaulting firm's capital. Banks receive the portion νk, while k_0 is transferred to the household and recycled to finance newly entering firms. This assumption ensures that entry is not fueled by exogenous resource injections, but rather by reallocating existing capital within the economy. Consequently, k_0 directly influences the recovery rate, alongside the parameter ν, which governs the share of physical capital recovered by lenders. Beyond its role in recovery, k_0 also affects firm leverage decisions: a smaller k_0 implies lower initial production capacity, requiring firms to borrow more if they aim to grow quickly. Banks internalize this demand for external financing and adjust contract terms accordingly.

Banks. I set the capacity-constraint parameter κ to 0.1, consistent with the ratio of the tenth-largest bank to the largest bank (see Figure C14 in Appendix C.4.2), in order to mimic the dominant-fringe market structure observed in the data. While this calibration anchors the model to an empirical reference point, Appendix B.4 shows that the stationary equilibrium depends only on the product $\mathcal{B}\kappa$, and similar aggregate outcomes can be obtained under alternative configurations that accommodate a longer tail of small banks.

The fixed entry cost F_E for big banks is chosen to match the aggregate profitability of the banking sector. For simplicity, I assume $F_E^s = \kappa \cdot F_E$. Given the small bank capacity constraint $l_b^{s\prime} = \kappa \cdot l_b^{\prime}$, the small bank value function is proportional: $V_b^s(X') = \kappa \cdot V_b(X')$.

[16] I also verify numerically that the firm equity value remains strictly positive for all firms in the distribution, hence introducing endogenous firm default in this baseline model would be immaterial under the current calibration. See Section 7 for an extension that incorporates endogenous firm default.

Under these assumptions, the entry conditions for big and small banks collapse to the same condition. Specifically, the small bank entry condition $\mathcal{M}(X, X') \cdot V_b^s(X') \geq F_E^s$ becomes $\mathcal{M}(X, X') \cdot V_b(X') \geq F_E^s/\kappa$, which is identical to the big bank entry condition. In stationary equilibrium, $X = X'$, with $B = B'$ and $\mathcal{B} = \mathcal{B}'$. To anchor the composition of entrants, I first determine the number of big banks B' such that adding one more big bank would render entry unprofitable, and then choose the number of small banks $\mathcal{B}'$ to finely adjust the degree of market competition and match aggregate profitability.

The aggregate profitability of the banking sector is calculated as *net operating income* over *total interest income*, which can be obtained from the *Quarterly Income and Expense of FDIC-Insured Commercial Banks and Savings Institutions*.[17] In particular, I use the average of a quarterly time series from 1984 to 2018. To summarize, I target: (i) the profitability of the banking sector through the fixed entry cost F_E, (ii) the frequency of equity issuance through the parameter λ_0, (iii) the market leverage through the parameter k_0, (iv) the recovery rate through the parameter ν, and (v) the standard deviation of investment rate through the parameter σ_z. The model targets a cross-sectional standard deviation of investment rates of 0.337, in line with other studies; see Khan and Thomas (2013), Clementi and Palazzo (2016), and Lanteri (2018).[18] All remaining moments are untargeted and reported for validation. Table I summarizes all targeted and untargeted moments.

Table I. Stationary Equilibrium and Annualized Moments

Targeted	**Description**	**Moment**	**Model**	**Data**
Yes	Profit/Revenue	$\pi_b / \int r_l(x, X) l_b d\Phi$	16.5%	16.2%
Yes	Freq. of Equity Iss.	$4 \int (d(x, X) < 0) d\Phi$	4.3%	4.2%
Yes	Market Leverage	$\int (Bl_b(x, X) + \mathcal{B}l_b^s(x, X))/V_F(x, X)\, \mathrm{d}\Phi$	38%	34%
Yes	Std. Investment Rate	$2\sqrt{\int (i/k)^2 d\Phi - \left(\int (i/k) d\Phi\right)^2}$	0.334	0.337
Yes	Recovery Rate	$\int \mathcal{I}(\nu k + k_0)/(\mathcal{I}Bl_b(x, X) + \mathcal{I}\mathcal{B}l_b^s(x, X)) d\Phi$	0.46	0.51
No	Capital to GDP	$K/(4Y)$	2.2	2.2
No	Investment to K	$4I/K$	14%	16%
No	Debt Adjust. to K	$\int 4(B\Delta l_b'(x, X) + \mathcal{B}\Delta l_b^{s\prime}(x, X))\, \mathrm{d}\Phi/K$	0.92%	0.62%
No	Num. Big Banks	B	3	-
No	Num. Small Banks	$\mathcal{B}$	7	-
No	Std. MPK	$2\sqrt{\int (Zz\alpha k^{\alpha-1})^2 d\Phi - \left(\int Zz\alpha k^{\alpha-1} d\Phi\right)^2}$	0.16	0.68

Notes: This table reports the targeted and untargeted aggregate annualized moments. Capital and investment in the data are computed from the current-cost net stock of private fixed assets, available from the U.S. Bureau of Economic Analysis. Loans in the data are computed from C&I loans from FRED. Leverage in the data is computed as leverage of non-financial corporate business, measured as debt as a percentage of the market value of corporate equities, also available from FRED. All moments refer to the average of the yearly time series from 1997 to 2017.

The calibrated model features an economy populated by 3 big and 7 small banks. More

[17] https://www.fdic.gov/analysis/quarterly-banking-profile/index.html.

[18] This is consistent with micro-level evidence from Cooper and Haltiwanger (2006).

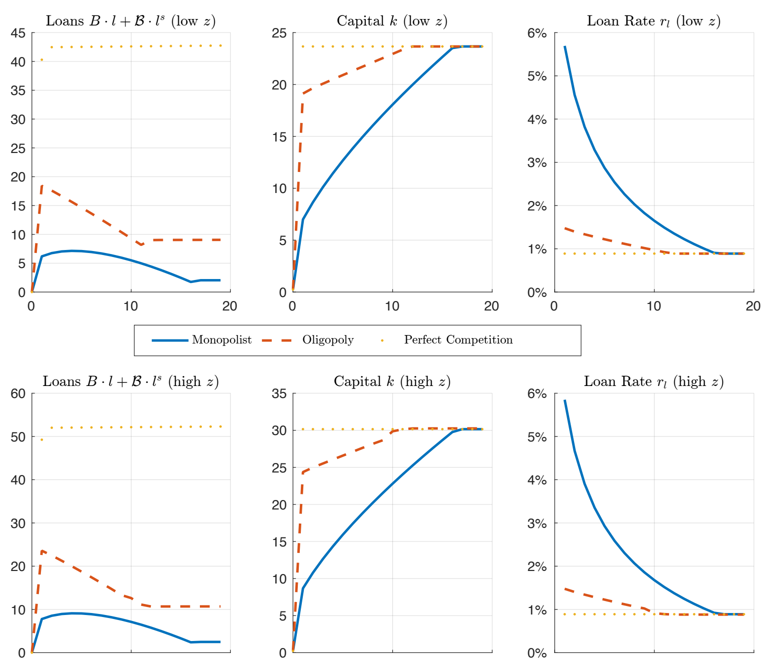
generally, I view the number of banks in the model as a proxy for the degree of strategic interaction in the market rather than to match the institutional count. See Appendix B.4 for further discussion. Note also that the banking sector in the U.S. is characterized by a dominant-fringe market structure. As an example, in 2018, the four largest banks in terms of assets were: (i) JPMorgan Chase & Co with a market share of 14.45%, (ii) Bank of America Corp. with a market share of 11.73%, (iii) Wells Fargo & Company with a market share of 11.17%, and (iv) Citigroup Inc. with a market share of 9.32%. The fifth bank was U.S. Bancorp with a significantly smaller market share of 3.02%, less than a third that of Citigroup Inc.

The value for the standard deviation of MPK, 0.68, is taken from David, Schmid, and Zeke (2022). My model generates, through bank market power alone, a value of 0.16 (∼23.5% of the total). I view the fact that the channel of bank market power can explain a relevant portion of the standard deviation of MPK but not all of it as natural, as there could be several other frictions at play as also described by David, Schmid, and Zeke (2022).

5.1 Firm Dynamics in the Stationary Equilibrium

I now describe the key properties of the stationary equilibrium of the calibrated model, with particular focus on the role of strategic interactions. Figure 2 reports the relevant stationary equilibrium policy functions across firms of different ages and productivity levels.

Figure 2. Stationary Equilibrium Policy Functions



Notes: This figure reports the stationary equilibrium policies for loan quantity (left panels), physical capital (middle panels), and loan interest rate (right panels), across firms of different ages and productivity levels. The x-axes indicate firm age.

Firms can reach their desired capital levels by (i) investing internal resources, (ii) issuing equity, or (iii) demanding external financing resources on the loan market. A more concentrated banking sector reduces credit availability in the economy. Firms with a high marginal productivity of capital and worse outside options, likely smaller or highly leveraged, exhibit a higher and less elastic credit demand. Therefore, these firms are more exposed to the negative effects of the lack of competition in the banking sector. This is the same intuition captured by equation (9) in the stylized model of Section 3. Since markups are endogenous in the cross-section of firms, banks endogenously exert a higher degree of market power on firms with a high marginal productivity of capital and lower internal resources; hence, fewer outside options but a higher marginal value from growth. These firms need bank credit and would otherwise incur an equity issuance cost to finance their growth, if their current production alone is not sufficient to sustain the desired level of physical investment. An imperfectly competitive financial sector internalizes that the same financial resources are more valuable for firms with a higher marginal productivity of capital and fewer outside options (e.g., a higher marginal equity issuance cost) and, therefore, can charge higher markups. This creates a mechanism of endogenous financial friction, as captured by the central panels of Figure 2.

Through this mechanism, limited competition in the financial sector not only induces credit misallocation—slowing firm growth—but also lowers aggregate productivity. This outcome is illustrated in Figure C1. The mechanism aligns with the empirical observation that firms more dependent on bank credit—such as small and private firms (Holmstrom and Tirole, 1997, Diamond and Rajan, 2005, Chodorow-Reich, 2013, Saunders, Spina, Steffen, and Streitz, 2024)—are disproportionately affected by limited competition, while firms with access to alternative funding sources, such as public bond markets, are less exposed to these frictions (Chava and Purnanandam, 2011).

As noted by Dempsey (2025), "while commercial banks account for a sizable subset of total lending in the U.S. economy, other lenders also supply significant amounts of credit." While the model can, in principle, be interpreted as referring to financial intermediaries more broadly, the quantitative calibration focuses specifically on bank lending. Incorporating non-bank financial institutions into the model could help capture substitution across credit channels and offer a more comprehensive picture of aggregate credit dynamics. Importantly, the mechanism of endogenous financial frictions described in this section would remain operative in such an environment. In connection with Figure 2, substitution toward non-bank finance appears more likely in the middle market segment, where relatively larger firms have better access to alternative sources of funding.

Figure C2 in Appendix C.1 reports the inverse elasticities $\eta_l'(x, X, x', X')$ from the generalized Euler equation (23), evaluated in the stationary equilibrium for firms of different ages. These elasticities, which are endogenous and depend on current and future firm-level characteristics, are directly linked to the equilibrium trajectories of markups. A lower inverse elasticity translates into a higher financial markup. Furthermore, Figure C2 in Appendix

C.1 shows that the higher the concentration of the banking sector, the lower the inverse elasticities and the longer it takes for a firm to reach its efficient level of capital. Under perfect competition, the inverse elasticity is constant in the cross-section of firms. Long-lived firms still experience a non-zero elasticity because of the interaction of bank market power with the curvature of the household utility function and the presence of a tax shield, so the level of loan is always well-defined.[19]

5.2 Validating the Mechanism

Qualitatively, this mechanism is consistent with stylized fact 3 ("controlling for deal amount and a proxy for corporate default, smaller firms tend to pay higher credit spreads"), with parameters of interest reported in Table C7 in Appendix C.4.1. This is qualitatively consistent with the model since, as suggested by the right panels of Figure 2, interest rates (hence, credit spreads) are higher for the smaller firms, holding default risk fixed.

Quantitatively, I compare the stationary equilibrium results from Section 5 with the empirical regression presented in Table C7. It is worth noting that this quantitative analysis should be interpreted with caution, as the Compustat-DealScan dataset includes public firms, which are inherently large, whereas the model suggests that smaller firms are the most affected by market power. Nevertheless, the variation in firm size within Compustat-DealScan can still be used to conduct the analysis.

As presented in Table C1, firms begin with a calibrated initial capital (k_0) of 0.237. In the calibrated oligopoly depicted in Figure 2, k_0 corresponds to a quarterly credit spread (relative to perfect competition) of approximately 0.587 percentage points.[20] Over time, firms grow to their efficient capital level and ultimately pay the competitive rate, implying that credit spreads relative to the perfect competition scenario converge to zero. In the calibrated oligopoly (illustrated by the dashed-red line), this occurs at a capital level of approximately 27.6, aggregating firms of varying productivity. Since the dependent variable in regression (1) of Table C7 is the all-in-drawn spread divided by 100, the coefficient on firm size is measured in percentage points of annualized credit spreads. Therefore, if default risk is held constant as in the model, regression (1) in Table C7 predicts the subsequent decline in the annualized credit spread:

$$\underbrace{-0.201 \cdot [\log(27.6) - \log(0.237)]}_{\simeq -0.96 \text{ p.p. annualized}} -0.198 \cdot [\log(9.8743) - \log(B \cdot l_0 + \mathcal{B} \cdot l_0^s)] - 0.157 \cdot \underbrace{0}_{\text{Const. Altman}} .$$

The second term is negative since, consistently with the model, firms start with zero loans.[21]

[19] See Appendix B.2 for further details.

[20] Under perfect competition, interest rates reflect the risk-free rate plus compensation for default risk, which is uniform across firms in the baseline model. Thus, credit spreads represent the portion not attributable solely to default risk and, within the model's framework, are interpreted as markups.

[21] Although an exact zero is not feasible in logarithmic terms, the reader can think of small initial values.

Hence, the second term contributes to a reduction in interest rates. Thus, focusing exclusively on the interest rate reduction due to changes in firm size represents a conservative estimate. For comparison with the model, I divide -0.96 percentage points by 4, resulting in a quarterly rate reduction of -0.24 percentage points. Within this range, the model predicts a quarterly credit spread reduction of -0.587 percentage points, demonstrating a comparable magnitude. The fact that the model yields a larger reduction aligns with the observation that the Compustat-DealScan dataset primarily consists of public firms, which are inherently larger and therefore less impacted by market power, consistent with the model's predictions.

Finally, to assess the sensitivity of credit spreads to market structure, I re-solve the stationary equilibrium after slightly reducing the capacity of small banks from $\kappa = 0.10$ to $\kappa = 0.095$, which increases the market share of large banks by about one percentage point. The resulting change in market structure raises the stationary-equilibrium annualized aggregate credit spread by roughly 2 basis points. The magnitude is comparable in sign and within the same order of magnitude as stylized fact 2, which estimates that a one-percentage-point increase in C_5 is associated with a 0.7% increase in the annualized loan spread. Using the average spread in the dataset used to compute stylized fact 2, a 0.7% increase in the annualized spread corresponds to roughly 3 basis points, suggesting that the model generates a moderately conservative and broadly consistent elasticity between bank concentration and credit spreads.

5.2.1 Relationship to the Empirical Literature

An extensive body of empirical research has explored the impact of bank competition on firms, with seminal works by Rajan and Zingales (1998), Black and Strahan (2002); Cetorelli and Gambera (2001); and Cetorelli and Strahan (2006). While evidence varies, the consensus aligns with my model, indicating that bank market power reduces the total credit available, though the effect differs across firms. Young firms face higher credit demand and are more affected by limited competition (see, Freixas and Rochet, 1997). The less competitive the credit market, the lower lenders' incentives to finance newcomers as documented by Petersen and Rajan (1995) and Cetorelli and Strahan (2006). Although my model does not focus on capturing complex features of relationship banking, the desire for intertemporal smoothing of bank profits and household consumption – embedded in the dynamic contract of equations (15), (16), and (23) – captures the fact that creditors, when setting markups, take into account firms' expected future dividend streams, as well as their own future profits. Hence, in my model, banks balance markups intertemporally in the spirit of relationship lending. This important economic force is consistent with Petersen and Rajan (1995) and Cetorelli and Strahan (2006).

6 Aggregate Shocks

This section analyzes the role that bank market power plays in the transmission of macroeconomic shocks. The section includes three unexpected aggregate shocks: (i) a decrease in aggregate TFP combined with an increase in the aggregate firm default probability, (ii) a temporary change in bank market structure (in the model, the first shock combined with an idiosyncratic shock to the assets of the big banks calibrated to push one bank to default), and (iii) a permanent change to the bank market structure (in the model, the second shock combined with a permanent increase in the fixed cost F_E to enter the bank market, calibrated to push one bank to default without a subsequent new bank reentry on the equilibrium path).

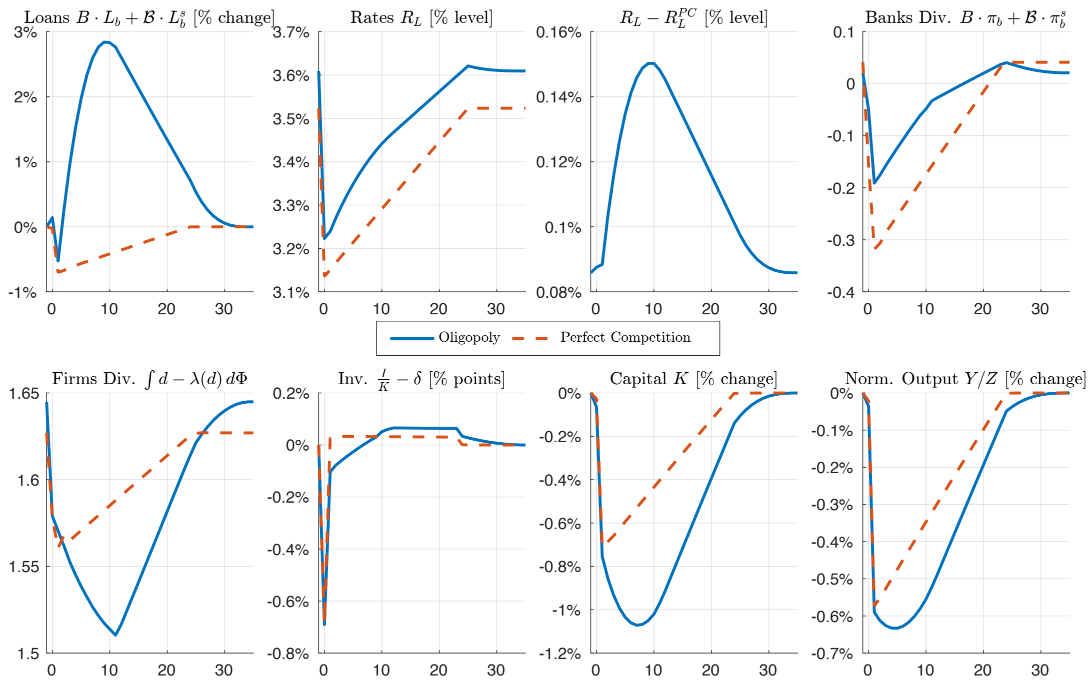
I compute the transitional dynamics of the model initialized at the stationary equilibrium as defined in Section 5. Then, I hit the economy with the unexpected aggregate shocks and, depending on the type of shock, the economy converges to the old (or a new) stationary equilibrium in the long run. Several papers assume agents did not foresee the aggregate shocks of the Great Recession (e.g., Guerrieri and Lorenzoni, 2017).

Along the transitional dynamics after the shock, I assume that all agents perfectly foresee the paths of all aggregate variables. To compute the equilibrium dynamics, I find sequences of: (i) aggregate consumption $\{C_t\}_{t=0}^T$; and (ii) firm distributions $\{\phi_t(x_t)\}_{t=0}^T$, such that households maximize utility, all markets clear in each period, and firm distributions evolve according to: (i) firm policy functions; (ii) the bank generalized Euler equations (23); and (iii) idiosyncratic default shocks.

6.1 A TFP Shock

This section investigates the effects of bank market power when the economy is hit at time $t = 0$ by an unexpected negative aggregate TFP shock of 2.5% (in correspondence of which the quarterly firm survival probability ρ decreases by 0.7%) as shown by Figure C4 in Appendix C.1. The shock is calibrated to match the size of the Great Recession.

Figure 3. TFP Shock, Financing, and Real Activity



Notes: This figure reports the transitional dynamics of: (i) the aggregate loans $B \int l_{b,t}(x_t, X_t)\, \mathrm{d}\Phi_t$, (ii) the annualized aggregate interest rate $\int r_{l,t}(x_t, X_t)\, \mathrm{d}\Phi_t$, (iii) the annualized aggregate interest rate spread calculated as the difference between the annualized aggregate interest rate $\int r_{l,t}(x_t, X_t)\, \mathrm{d}\Phi_t$ under the calibrated oligopoly and the annualized aggregate interest rate $\int r_{l,t}^{\mathrm{PC}}(x_t, X_t)\, \mathrm{d}\Phi_t^{\mathrm{PC}}$ under perfect competition, (iv) the aggregate bank dividend $\sum_{b=1}^{B_t} \pi_{b,t}$, (v) the annualized aggregate firm dividend $\int d_t(x_t, X_t) - \lambda(d_t(x_t, X_t))\, \mathrm{d}\Phi_t$, (vi) the aggregate physical investment $(\int i_t(x_t, X_t)\, \mathrm{d}\Phi_t)/(\int k_t\, \mathrm{d}\Phi_t)$ net of depreciation (quarterly), (vii) the aggregate capital $\int k_t\, \mathrm{d}\Phi_t$ and (viii) the aggregate output (normalized to remove the exogenous component Z_t to isolate the effects on the endogenous component) $\int k_t^{\alpha}(x_t, X_t)\, \mathrm{d}\Phi_t$, following the TFP shock reported in Figure C4 in Appendix C.1. The x-axes report time t.

Note that the bank entry condition (30) and the bank exit condition (19) must hold during the transition. Given the calibration of Table I, this shock does not induce any new bank to enter or an incumbent bank to exit. Figure 3 reports the dynamic responses calculated (i) with the calibrated oligopolistic banking sector of Section 5 (solid line) and (ii) with the corresponding perfectly competitive banking sector (dashed line). As discussed in Subsection 5.1, a more concentrated banking sector can extract higher rents out of the financially constrained firms with a high marginal productivity of capital. In the stationary equilibrium, this mechanism endogenously creates slower growth trajectories as a function of the bank market structure. In the dynamics, this mechanism of endogenous financial frictions interacts with the higher density of financially constrained firms, generating a higher credit demand that drives up markups as shown in Figure 3.[22] Therefore, during the transitional dynam-

[22]This is consistent with the idea that firms more dependent on external funding via bank loans, such as small and private firms, can become more financially constrained when credit conditions tighten (Holmstrom

ics, interest rates rise more under oligopolistic competition than under perfect competition. Moreover, total loans rise under oligopolistic competition. Banks exploit their market power not only to extract higher markups, but also to prevent their profits from falling as much as under perfect competition. When the shock hits, banks incur large losses and require an equity injection from households (i.e., a negative bank dividend). Nevertheless, the shock is calibrated to ensure that no bank defaults in equilibrium, as $\forall b = \{1, \ldots, B\} : \tilde{V}_{b,0} > 0$ and $\forall b = \{1, \ldots, \mathcal{B}\} : \tilde{V}^s_{b,0} > 0$, where $\tilde{V}_{b,0}$ and $\tilde{V}^s_{b,0}$ denote the time-0 value functions of large and small banks, respectively.

In summary, in such conditions, a concentrated banking sector exploits its market power to extract higher interest rates. This mechanism induces a larger decline in real activity in terms of aggregate investment, capital, and output. Finally, note that the decline in real activity captured by Figure 3 further constrains firms by limiting the household's capacity to support firms with equity issuance (the aggregate firm dividend declines). This mechanism creates a vicious cycle that further reduces the interest rate-quantity loan elasticities $\eta_l'(x, X, x', X')$ and raises bank loan rates. Throughout the paper, I assume that the mass of defaulting firms is replaced, at each date t, by an equal mass of new entrant firms. In Subsection 6.2, I examine the effects of an entry rate temporarily below the exit rate, as occurred during the Great Recession.

6.2 The Effects of the Firm Entry Rate

During the Great Recession, the firm default rate increases, but not all firms immediately reenter the market. In this section, I let the mass of exiting firms evolve according to the evolution of the default rate in the shock reported in Figure C4 in Appendix C.1. Differently from before, I keep the entering mass of firms initially lower than the exit mass of firms and higher after 10 periods. The entry rate, along the shock, is calibrated so that the mass of firms in the production sector drops in the short-run by about 1 percent and returns to mass 1 in the long-run. In Figure C3 in Appendix C.1, I refer to this type of shock with the label *Variable Entry Mass*. In contrast, the label *Entry Mass = Exit Mass* refers to the shock already analyzed in the previous Subsection 6.1, whose effects on real activity are reported in Figure 3. Figure C3 suggests that, when the firm default rate increases but not all firms immediately reenter the market, then imperfect competition in the financial intermediation sector leads to a larger and delayed amplification effect at the peak that fades away as new firms enter the production sector. The intuition is linked to the economic mechanism discussed in Subsection 6.1. Banks can extract higher interest rates for longer as firms slowly reenter the production sector (as shown by the left panel of Figure C3). A concentrated banking sector extracts higher rents out of the financially constrained firms that slowly reenter the market and need bank credit. Output eventually converges back to

and Tirole, 1997; Diamond and Rajan, 2005; Chodorow-Reich, 2013; Saunders, Spina, Steffen, and Streitz, 2024).

its pre-shock level as the mass of producing firms returns to one.

6.3 Changes in Bank Market Structure

One salient ingredient of my framework is the presence of non-atomistic financial intermediaries. Thanks to this feature, I can study market structure changes, such as bank failure and bank reentry. The economy is hit at time $t = 0$ by the same shock of Subsection 6.1, which is reported in Figure C4 in Appendix C.1. Furthermore, I introduce an idiosyncratic shock ζ_t^χ to the assets of the banks such that equation (20) becomes:

$$\pi_b^\chi = (1 - \zeta_t^\chi) \cdot \int \mathcal{I} \cdot r_l \cdot l_b^\chi \,\mathrm{d}j + r_M M_b^\chi - r_D D_b^\chi - F^\chi. \tag{32}$$

I calibrate the shock so that it is rational for one big bank to default on the equilibrium path at time $t = 0$. In particular, the shock is such that $\forall t : \zeta_t^s = 0$ (small banks) and $\zeta_0 = 0.016$ (big banks) and then it reverts back to 0 linearly with a persistence calibrated consistently with that of the TFP shock. In Subsection 6.4, I report the resulting equilibrium path which includes the entry of a new bank at time $t = 21$. This is in line with the events of the Great Recession, during which there were numerous bank failures, bailouts, as well as significant failures and mergers involving major banks (e.g., the bankruptcy of Lehman Brothers, the collapse of Washington Mutual, and the acquisition of Wachovia by Wells Fargo in 2008). Additionally, the Great Recession was marked by a general trend of consolidation in the banking market structure, accompanied by a sharp decline in new bank entries since 2009. In order to mimic the collapse in bank entry and the long-run trend of market structure consolidation of the banking sector, in Subsection 6.5, I combine the shock of Subsection 6.4 with a permanent increase in the fixed cost of entry F_E, so that it is not profitable for a new bank to enter the credit market along the equilibrium path.

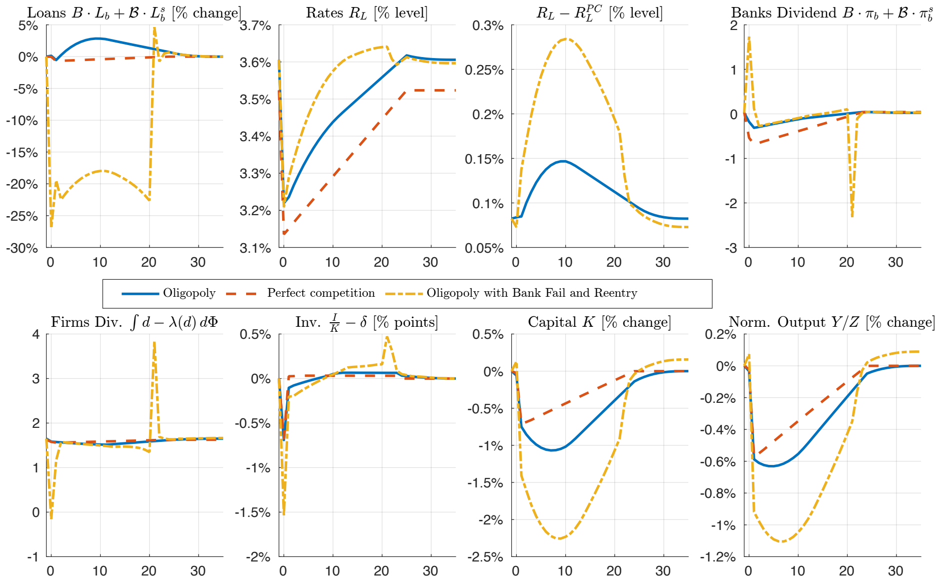
6.4 Bank Failure and Bank Market Power

In this subsection, the fixed cost of entry F_E is held fixed throughout the dynamics, allowing one bank to enter the credit market along the equilibrium path. At $t = 0$, when the shock hits, one incumbent big bank decides to exit since $\tilde{V}_{b,0} < 0$ and the market structure changes temporarily from 3 big banks and 7 small banks to 2 big banks and 7 small banks. Note that the shock is calibrated such that exactly one big bank chooses to default; that is, when one bank exits at $t = 0$, the value functions of all remaining banks become positive.[23] Additionally, since all large banks are identical, any of them could potentially default. It is immaterial for the equilibrium dynamics which specific bank defaults because, due to symmetry, any default scenario would yield identical equilibrium outcomes. The equilibrium path is deter-

[23] This requires solving for the unexpected shock multiple times and checking the resulting value functions under different market structures.

mined by the fact that, given it is optimal for exactly one bank to default—regardless of which one—it is also optimal for the remaining banks to continue operating.

Figure 4. Bank Default and Reentry, Financing, and Real Activity



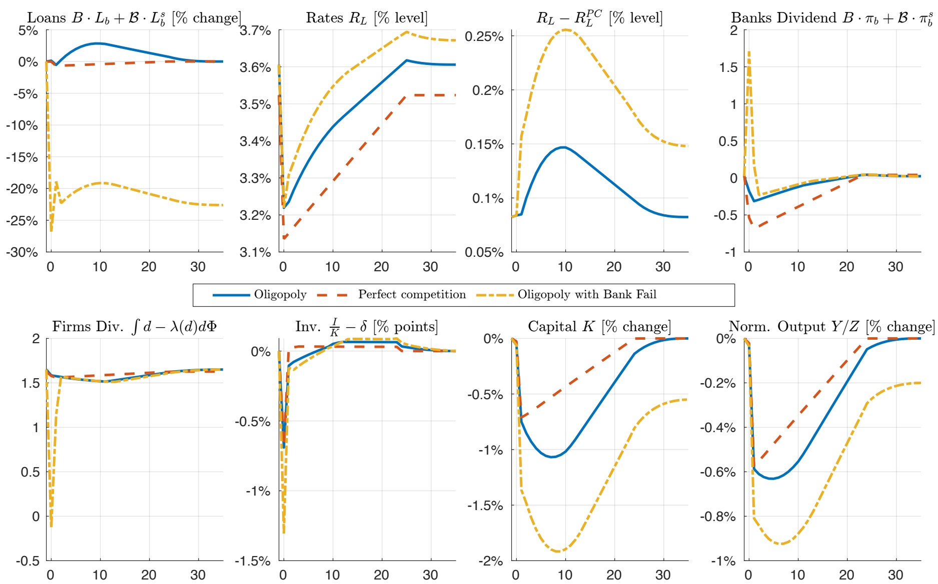
Notes: This figure reports the transitional dynamics following the TFP shock reported in Figure C4 in Appendix C.1. At $t = 0$ (on impact) one incumbent big bank exits. At $t = 21$, it is profitable for one big bank to enter the market given that the fixed cost F_E is held fixed at its stationary equilibrium value. The x-axes report time t.

At $t = 21$, it is profitable for a new bank to enter the credit market because the fixed cost F_E has not changed. Hence, the market structure changes back from 2 to 3 big banks. Figure 4 reports the resulting dynamics. At the beginning of period $t = 0$, when one bank decides to exit, non-defaulting firms fully repay their outstanding loans to all banks, including the exiting bank. To prevent the destruction of assets, the defaulting bank liquidates its remaining assets, distributing them to the household as dividends. All loans are repaid prior to the change in market structure, as agents only trade one-period securities. Following the change, all agents make decisions in accordance with the new market structure. When a big bank fails, the surviving banks slowly begin extending more credit to firms in order to partially capture the market share of the defaulted bank. However, the speed of this adjustment is dampened by the decreased level of competition among surviving banks. The surviving banks' market power interacts with credit constraints, yielding a sharp drop in the aggregate volume of credit in the short run with interest rates that quickly increase after the initial drop. Moreover, banks reduce credit supply in anticipation of the new bank's entry. The credit crunch propagates into the real economy, yielding a sharp and persistent drop in investment; hence physical capital and output. The entry of the new bank at $t = 21$ requires an initial equity injection from households (i.e., a negative bank dividend). The resulting expansion in credit supply increases loan availability and lowers interest rates, which raises the aggregate firm dividends. Investment in physical capital also temporarily rises, bringing capital and output back to their pre-default levels.

6.5 A Permanent Change in Bank Market Structure

In this subsection, I combine the shock of Subsection 6.4 with a permanent increase in the fixed entry cost F_E, so that it is not profitable for a new bank to enter the credit market along the equilibrium path.

Figure 5. Bank Default, Financing, and Real Activity



Notes: This figure reports the transitional dynamics following the TFP shock reported in Figure C4 in Appendix C.1. At $t = 0$ (on impact) one incumbent big bank exits. The x-axes report time t.

When the market structure of the banking sector changes permanently, the model captures two ideas: (i) in the short term, the market power of the surviving banks contributes to lowering the supply of credit to firms, further slowing down the economy and (ii) in the long run, the heightened bank market power further contributes to amplifying and prolonging the recession. Figure 5 reports the resulting dynamics. Because of the general equilibrium effects and the reduced level of competition, in the long run, the economy stabilizes at a lower volume of credit, which results in less investment, capital, and output. Subsection 6.6 discusses the dynamic effects on the dispersion of loan rates (associated with the dispersion of marginal products of capital) and aggregate TFP.

6.6 Dispersion of Loan Rates and Aggregate TFP

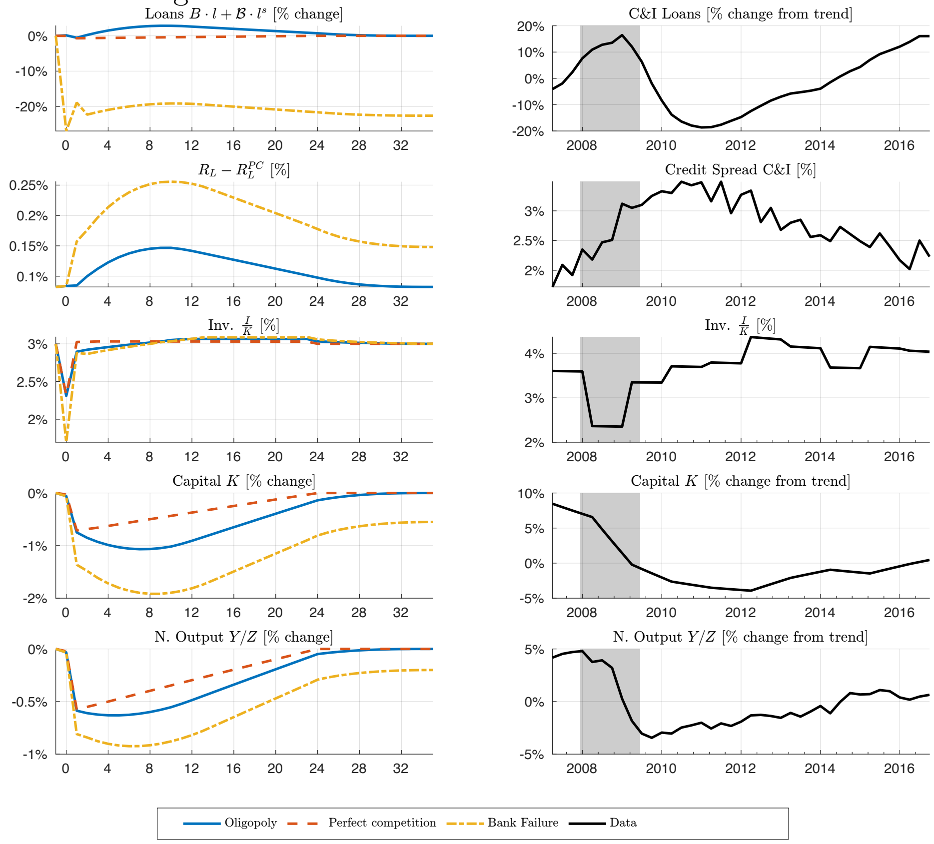
In this subsection, I study the effects on the dispersion of loan rates and aggregate TFP of the shock of Subsection 6.5. The dynamic financial oligopoly combined with heterogeneous firms generates firm-level endogenous financial frictions that create time-varying second moments, such as the dispersion of loan rates (directly linked to the dispersion of marginal products of capital) and aggregate TFP. The left panel of Figure C8 in Appendix C.1 reports the dynamics of the standard deviation of loan rates, expressed in percentage levels. The right panel of Figure C8 reports the associated aggregate TFP (calculated as the residual of an aggregate production $Y_t = \text{TFP}_t \cdot K_t^{\alpha}$). Both measures are calculated as difference from the perfectly competitive benchmark.

In agreement with empirical evidence (e.g., Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry, 2018 and David, Schmid, and Zeke, 2022), the model produces dynamics with an increasing dispersion of loan rates during recessions; hence, an increasing dispersion of marginal productivity of capital. Figure C8 also suggests that the failure of a large bank induces a small but persistent decline in aggregate TFP. This result highlights that oligopolistic banks amplify business cycles primarily through higher spreads and reduced credit demand, rather than through quantitatively meaningful misallocation across firms. The contrast between the importance of this misallocation in the stationary equilibrium and in dynamics is also worth noting. The failure of a large bank induces a permanent decline in aggregate TFP of a few basis points—quantitatively limited for business cycles but potentially relevant for welfare.

6.7 Comparison with the Great Recession

Figure 6 shows that the model dynamics of Subsection 6.5 are broadly consistent with those of the Great Recession. For example, in the aftermath of the financial crisis, C&I credit spreads increased at the peak by 1.5 percentage points (annualized). As shown in the second panel on the left, the model suggests that, at the peak, approximately 0.15 percentage points and 0.25 percentage points of these credit spreads are attributable to financial markups (calculated with and without bank default, respectively). Moreover, the model captures a drop in investment rate similar in magnitude to that of the Great Recession, and a persistent drop in capital (around 2% at the peak) and output (around 1% at the peak).

Figure 6. Model vs. the Great Recession



Notes: This figure compares the model dynamics of Subsection 6.5 with the data around the Great Recession. The left column reports model responses in quarters. The right column reports the corresponding data: i) C&I loans, expressed in percent deviation from a linear trend estimated over 1997:Q2–2017:Q2; ii) C&I credit spreads; iii) the quarterly investment rate; iv) physical capital, expressed in percent deviation from trend; and v) normalized output, also expressed in percent deviation from trend.

The resulting increase in bank market power amplifies and prolongs the recession, suggesting that bank market power may have played a significant role in amplifying and prolonging the crisis. The figure also depicts the dynamics of total loans. Around the onset of the Great Recession, total loans decline in the model following a bank failure, in line with the empirical evidence. Without bank reentry, total loans remain persistently depressed as the economy converges to a permanently lower level of total credit. In contrast, as shown by Figure C6 in Appendix C.1, when endogenous bank reentry occurs along the equilibrium path, banks reduce lending by less in response to a failure—anticipating future reentry—and the reentry itself restores total loans to their pre-crisis steady state. In this sense, bank reentry helps generate total lending dynamics that more closely align with those observed in the data.

7 Bank Market Power and Endogenous Firm Default

In this section, I extend the model to allow bank market power to interact with firm default decisions. Banks strategically internalize the effects of their choices on both current and future firm defaults, while also interacting strategically with other banks. This problem is particularly challenging, as it features a non-differentiable dynamic demand for loans, which renders the Markov perfect equilibrium non-differentiable and precludes the use of the generalized Euler equations approach. I address this challenge using dynamic discrete choice methods, which yield smooth decision rules and restore differentiability. Similar methods have long been used in various areas of economics (e.g., McFadden, 1977), including industrial organization and, more recently, the sovereign default literature.[24] Following this approach, I introduce shocks to firm outside option values, assumed to follow a logistic distribution. I then derive the generalized Euler equation that captures the interaction between bank market power and firm default behavior.

7.1 Firms

The value function $V_F(x, X)$ of a firm satisfies the following functional equation:

$$\tilde{V}_F(x, X) \;=\; \max_{\{l_b'\}_b^{\mathcal{B}'},\{l_b^{s\prime}\}_b^{\mathcal{B}'},k'} \; d - \lambda(d) + \mathbb{E}\left[\mathcal{I}' \cdot \mathcal{M}' \cdot V_F(x', X')|(x, X)\right], \tag{33}$$

where, at the beginning of each period, each firm decides whether to default or not $\mathcal{I}^d = \{0, 1\}$, so that the function $V_F(x, X)$ satisfies the following functional equation

$$V_F(x, X) \;= \max\left\{\tilde{V}_F(x, X), \epsilon\right\} = \max_{\mathcal{I}^d=\{0,1\}} \; \mathcal{I}^d \cdot \tilde{V}_F(x, X) + (1 - \mathcal{I}^d) \cdot \epsilon, \tag{34}$$

where the outside value shock ϵ follows a logistic distribution with mean zero and scale parameter $\zeta\,\bar{V}_F$, and $\bar{V}_F > 0$ is the average firm equity value in the stationary equilibrium used for normalization. All firm constraints are identical to those specified in Section 4.

Use equation (34) to evaluate the expectation of the value function with respect to the outside value shocks:[25]

$$\mathbb{E}^{\epsilon}[V_F(x, X)] = \tilde{V}_F(x, X) + \zeta\bar{V}_F \cdot \ln\left(1 + e^{-\zeta^{-1}\frac{\tilde{V}_F(x,X)}{\bar{V}_F}}\right). \tag{35}$$

[24] Recent relevant contributions include Chatterjee and Eyigungor (2012); Dvorkin, Sánchez, Sapriza, and Yurdagul (2021); and Chatterjee, Corbae, Dempsey, and Ríos-Rull (2023).

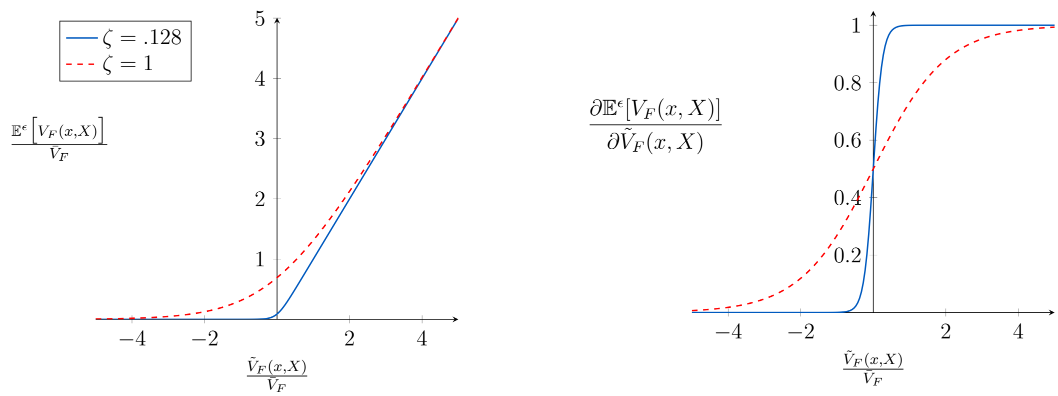
[25] See Appendix B.3 for mathematical details.

From equation (35) it follows that

$$\frac{\partial \mathbb{E}^{\epsilon}[V_F(x,X)]}{\partial \tilde{V}_F(x,X)} = \frac{1}{1+e^{-\zeta^{-1}\frac{\tilde{V}_F(x,X)}{\bar{V}_F}}} \equiv \tilde{\mathcal{I}}^d(\tilde{V}_F(x,X)), \tag{36}$$

is the logistic function which, intuitively, corresponds to the probability of a firm not defaulting ($\Pr[\mathcal{I}^d = 1|(x,X)]$) given all possible realizations of the exit value shocks.

Figure 7. Firm Expected Value with Logistic Shocks



Notes: This figure reports the firm's expected value and its derivative for two values of ζ.

Figure 7 illustrates the intuition behind equations (35) and (36). It shows how this smooth, differentiable formulation nests the textbook case—where shareholders' outside option is deterministically zero—and converges to it as the scale parameter ζ approaches zero. I provide a formal proof in Proposition II presented in Appendix B.3. The left panel plots the normalized expected equity value $\mathbb{E}^{\varepsilon}\left[V_F(x,X)\right]/\bar{V}_F$ from (35) against $\tilde{V}_F(x,X)/\bar{V}_F$. The right panel shows the corresponding solvency probability $\tilde{\mathcal{I}}^d$, i.e. the logistic CDF, as a function of $\tilde{V}_F(x,X)/\bar{V}_F$. In both panels the solid curve is $\zeta = .128$ (the calibrated value as described in the next subsection) and the dashed curve is $\zeta = 1$. Thanks to its "softmax" form, the mapping remains everywhere differentiable but collapses to the textbook indicator as $\zeta \to 0^+$. Equation (36) is useful for computing envelope conditions. The first-order condition with respect to k' is:

$$1-\lambda_d(d) = \mathbb{E}\left[\mathcal{M}'(X,X')\cdot\mathcal{I}'\cdot\tilde{\mathcal{I}}^d(\tilde{V}_F(x',X'))\cdot\Big(1+(1-\tau)(Z'z'\alpha k'^{\alpha-1}-\delta)\Big)\cdot(1-\lambda_d(d'))|(x,X)\right]. \tag{37}$$

The first-order condition with respect to $l_b^{\chi\prime}$ is:

$$1-\lambda_d(d) = \mathbb{E}\left[\mathcal{M}'(X,X')\cdot\mathcal{I}'\cdot\tilde{\mathcal{I}}^d(\tilde{V}_F(x',X'))\cdot(1+(1-\tau)r_l')\cdot(1-\lambda_d(d'))|(x,X)\right]. \tag{38}$$

Expressions (37) and (38) are key to deriving generalized Euler equations which, in turn, are essential to characterize bank behavior in a computationally tractable way.

7.2 Banks

The decision problem for a new entrant is identical to that in the baseline model. The decision problem for an incumbent bank now leads to modified generalized Euler equations that take into account how bank decisions affect firm default decisions. Similarly to the baseline model, given other banks' contracts $\{D'_{-b}, r'_D\}$ and $\{l'_{-b}(x,X), {l^s_{-b}}'(x,X), r'_l(x,X)\}$, a bank b best responds with a contract $\{D'_b, r'_D\}, \{l'_b(x,X), {l^s_b}'(x,X), r'_l(x,X)\}$ that satisfies the functional equations (18) and (19) subject to: (i) equation (17), (ii) the household's interest rate-quantity schedule jointly defined by equations (13) and (14), (iii) each firm's interest rate-quantity schedule which, in the extension, is jointly defined by equations (37) and (38). Each big bank's best response function satisfies the following generalized Euler equation:

$$1 = \mathbb{E}\left[\mathcal{I}' \cdot \mathcal{M}' \cdot \left(\mathcal{D}(x,X,x',X') + \tilde{\mathcal{I}}^d(x',X') \cdot \left(1 + \frac{l'_b}{R'_l}\frac{\partial R'_l}{\partial l'_b}\right)\right) \cdot R'_l|(x,X)\right]. \tag{39}$$

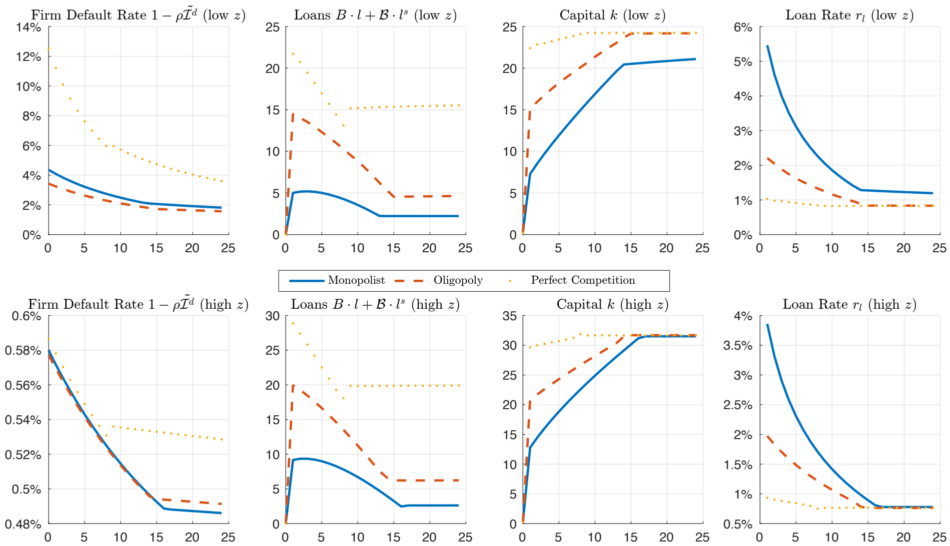
This equation further generalizes equation (23). For the sake of clarity, equation (39) is reported without the marginal recovery rate, which can be added similarly to equation (23). Also, a similar GEE arises for small banks, in the spirit of equation (24). The inverse elasticity can be computed using steps similar to those in Subsection 4.4.2. When firm default decisions are endogenous, a bank internalizes the impact that an additional unit of loan has on each firm's default decision according to

$$\forall z': \quad \mathcal{D}(x,X,x',X') \equiv \frac{\mathrm{d}\tilde{\mathcal{I}}^d(\tilde{V}'_F)}{\mathrm{d}\tilde{V}'_F} \cdot \left[\tilde{V}_{F,l_b}(x',X') \cdot l'_b + \tilde{V}_{F,R_l}(x',X')\frac{\partial R'_l}{\partial l'_b} \cdot l'_b\right].$$

Note that when firm default is treated exogenously, the term $\frac{\mathrm{d}\tilde{\mathcal{I}}^d(\tilde{V}'_F)}{\mathrm{d}\tilde{V}'_F}$ is zero and equation (39) collapses to equation (23). The term $\mathcal{D}(.)$ captures the strategic behavior of a bank that internalizes the marginal effect that an additional unit of loans has on the firm's future equity value and future default decisions. I set $\bar{\rho} = 1 - 0.21\%/2$ to maintain an effective level of default shocks in the stationary equilibrium. Hence, the new relevant parameter to calibrate is ζ, which I use to target an annual default rate of 0.84%.

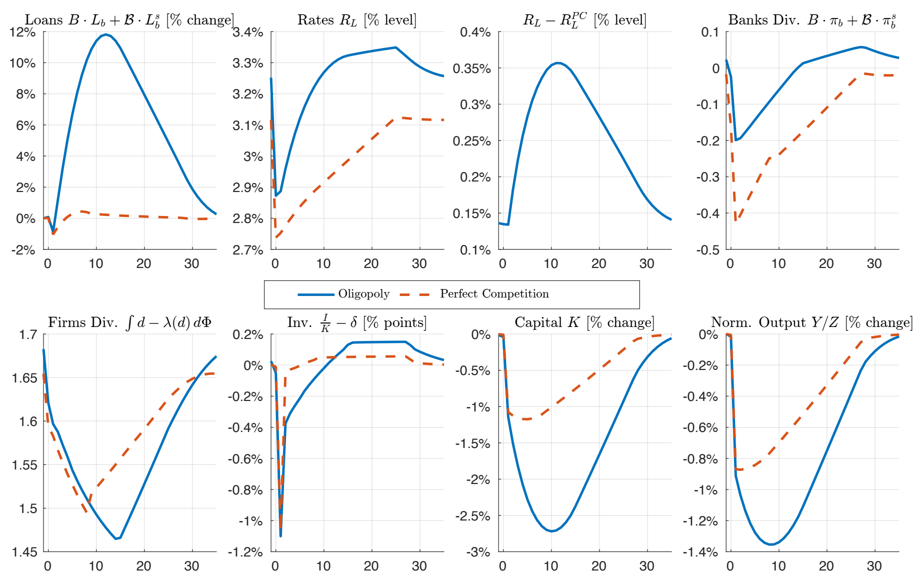
First, I calibrate the stationary equilibrium to target: (i) the profitability of the banking sector through the fixed entry cost F_E, (ii) the frequency of equity issuance through the parameter λ_0, (iii) the market leverage through the parameter k_0, (iv) the recovery rate through the parameter ν, (v) the standard deviation of investment rate through the parameter σ_z, and (vi) the default rate through the standard deviation of the outside value shocks ζ. All parameters resulting from the calibration process are reported in Table C2 in Appendix C.1, where the new parameter ζ is calibrated to be 0.128. All the corresponding aggregate stationary equilibrium moments are reported in Table C3 in Appendix C.1.

Figure 8. Stationary Equilibrium Policy Functions



Notes: This figure reports the stationary equilibrium policy functions when firms can endogenously default (the counterpart in the baseline model is Figure 2). The x-axes report the firm age.

Figure 9. TFP Shock, Financing, and Real Activity



Notes: This figure reports the transitional dynamics when firms can endogenously default (the counterpart in the baseline model is Figure 3), following the TFP shock reported in Figure C4 in Appendix C.1. The x-axes report time t.

Figure 8 reports the stationary equilibrium policy functions, which are qualitatively similar to those of the baseline model. Additionally, as shown by Figure 8, firms that are more financially constrained and more exposed to bank market power also face higher endogenous default risk. Second, I hit the economy with the same unexpected shocks of Subsection 6.5

and compare the dynamics. Figure 9 reports the transitional dynamics when firms can endogenously default following the TFP shock reported in Figure C4 in Appendix C.1. Figure C12 in Appendix C.3 complements Figure 9 and also reports the endogenous response of the aggregate firm default rate along the shock. The resulting dynamics are qualitatively comparable to the baseline model. Quantitatively, the effects of bank market power are amplified.

8 Conclusion

This paper develops a macroeconomic model that examines the role of bank market power in firm dynamics, motivated by rising concentration and markups in the U.S. banking sector. Incorporating oligopolistic banks and firm heterogeneity, the model formalizes how bank market power varies across firms with different characteristics, impacting firm growth, aggregate productivity, and output. The model shows that limited competition enables banks to charge firm-specific markups, disproportionately affecting productive financially constrained firms and dampening growth. Bank market power amplifies the effects of macroeconomic shocks, as banks extract higher markups during crises, thereby dampening capital accumulation and intensifying declines in real activity. Following a major bank's failure, surviving banks use their increased market power to restrict credit supply, deepening the economic downturn. These findings suggest that policymakers should consider the interaction between bank market power and credit constraints as an additional rationale for supporting large banks, beyond traditional concerns like preventing bank runs.

Data Availability Statement. The data and code underlying this article are available in Zenodo at https://doi.org/10.5281/zenodo.18527736.

References

Asplund, M., and V. Nocke (2006): "Firm Turnover in Imperfectly Competitive Markets," *The Review of Economic Studies*, 73(2), 295–327.

Bernanke, B., and M. Gertler (1989): "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review*, 79(1), 14–31.

Black, S. E., and P. E. Strahan (2002): "Entrepreneurship and Bank Credit Availability," *The Journal of Finance*, 57(6), 2807–2833.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. J. Terry (2018): "Really Uncertain Business Cycles," *Econometrica*, 86(3), 1031–1065.

Cetorelli, N., and M. Gambera (2001): "Banking Market Structure, Financial Dependence and Growth: International Evidence from Industry Data," *The Journal of Finance*, 56(2), 617–648.

Cetorelli, N., and P. E. Strahan (2006): "Finance as a Barrier to Entry: Bank Competition and Industry Structure in Local U.S. Markets," *The Journal of Finance*, 61(1), 437–461.

Chatterjee, S., D. Corbae, K. Dempsey, and J.-V. Ríos-Rull (2023): "A Quantitative Theory of the Credit Score," *Econometrica*, 91(5), 1803–1840.

Chatterjee, S., and B. Eyigungor (2012): "Maturity, Indebtedness, and Default Risk," *American Economic Review*, 102(6), 2674–99.

Chava, S., and A. Purnanandam (2011): "The effect of banking crisis on bank-dependent borrowers," *Journal of Financial Economics*, 99(1), 116–135.

Chava, S., and M. R. Roberts (2008): "How Does Financing Impact Investment? The Role of Debt Covenants," *The Journal of Finance*, 63(5), 2085–2121.

Chodorow-Reich, G. (2013): " The Employment Effects of Credit Market Disruptions: Firm-level Evidence from the 2008–9 Financial Crisis *," *The Quarterly Journal of Economics*, 129(1), 1–59.

Clementi, G. L., and B. Palazzo (2016): "Entry, Exit, Firm Dynamics, and Aggregate Fluctuations," *American Economic Journal: Macroeconomics*, 8(3), 1–41.

Clymo, A., and A. Lanteri (2020): "Fiscal Policy with Limited-Time Commitment," *The Economic Journal*, 130(627), 623–652.

Cooley, T. F., and V. Quadrini (2001): "Financial Markets and Firm Dynamics," *American Economic Review*, 91(5), 1286–1310.

Cooper, R., and J. Haltiwanger (2006): "On the Nature of Capital Adjustment Costs," *The Review of Economic Studies*, 73(3), 611–633.

Corbae, D., and P. D'Erasmo (2021): "Capital Buffers in a Quantitative Model of Banking Industry Dynamics," *Econometrica*, 89(6), 2975–3023.

Covas, F., and W. J. den Haan (2011): "The Cyclical Behavior of Debt and Equity Finance," *American Economic Review*, 101(2), 877–99.

David, J. M., L. Schmid, and D. Zeke (2022): "Risk-adjusted capital allocation and misallocation," *Journal of Financial Economics*, 145(3), 684–705.

De Loecker, J., J. Eeckhout, and G. Unger (2020): "The Rise of Market Power and the Macroeconomic Implications*," *The Quarterly Journal of Economics*, 135(2), 561–644.

Demirgüç-Kunt, A., and M. S. Martinez Peria (2010): "A Framework for Analyzing Competition in the Banking Sector: An Application to the Case of Jordan," *World Bank Policy Research Working Paper No. 5499.*

Dempsey, K. (2025): "Capital Requirements with Non-Bank Finance," *The Review of Economic Studies*, Forthcoming.

Diamond, D. W., and R. G. Rajan (2005): "Liquidity Shortages and Banking Crises," *The Journal of Finance*, 60(2), 615–647.

Dvorkin, M., J. M. Sánchez, H. Sapriza, and E. Yurdagul (2021): "Sovereign Debt Restructurings," *American Economic Journal: Macroeconomics*, 13(2), 26–77.

Elenev, V., T. Landvoigt, and S. Van Nieuwerburgh (2021): "A Macroeconomic Model With Financially Constrained Producers and Intermediaries," *Econometrica*, 89(3), 1361–1418.

Freixas, X., and J. Rochet (1997): "Microeconomics of Banking, Volume 1," *The MIT Press, 1 edition.*

Gomes, J. F. (2001): "Financing Investment," *American Economic Review*, 91(5), 1263–1285.

Guerrieri, V., and G. Lorenzoni (2017): "Credit Crises, Precautionary Savings, and the Liquidity Trap*," *The Quarterly Journal of Economics*, 132(3), 1427–1467.

Hennessy, C. A., and T. M. Whited (2005): "Debt Dynamics," *The Journal of Finance*, 60(3), 1129–1165.

——— (2007): "How Costly Is External Financing? Evidence from a Structural Estimation," *The Journal of Finance*, 62(4), 1705–1745.

Holmstrom, B., and J. Tirole (1997): "Financial Intermediation, Loanable Funds, and the Real Sector," *The Quarterly Journal of Economics*, 112(3), 663–691.

Jamilov, R., and T. Monacelli (2025): "Bewley Banks," *The Review of Economic Studies*, Forthcoming.

Jermann, U., and V. Quadrini (2012): "Macroeconomic Effects of Financial Shocks," *American Economic Review*, 102(1), 238–71.

Khan, A., and J. K. Thomas (2013): "Credit Shocks and Aggregate Fluctuations in an Economy with Production Heterogeneity," *Journal of Political Economy*, 121(6), 1055–1107.

Klein, P., P. Krusell, and J.-V. Ríos-Rull (2008): "Time-Consistent Public Policy," *The Review of Economic Studies*, 75(3), 789–808.

Klein, P., and J.-V. Ríos-Rull (2003): "Time-Consistent Optimal Fiscal Policy," *International Economic Review*, 44(4), 1217–1245.

Kreps, D. M., and J. A. Scheinkman (1983): "Quantity Precommitment and Bertrand Competition Yield Cournot Outcomes," *The Bell Journal of Economics*, 14(2), 326–337.

Lanteri, A. (2018): "The Market for Used Capital: Endogenous Irreversibility and Reallocation over the Business Cycle," *American Economic Review*, 108(9), 2383–2419.

Maskin, E., and J. Tirole (2001): "Markov Perfect Equilibrium: I. Observable Actions," *Journal of Economic Theory*, 100(2), 191–219.

McFadden, D. (1977): "Modelling the Choice of Residential Location," *Transportation Research Record*.

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

Midrigan, V., and D. Y. Xu (2014): "Finance and Misallocation: Evidence from Plant-Level Data," *American Economic Review*, 104(2), 422–58.

Petersen, M. A., and R. G. Rajan (1995): "The Effect of Credit Market Competition on Lending Relationships," *The Quarterly Journal of Economics*, 110(2), 407–443.

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

Saunders, A., A. Spina, S. Steffen, and D. Streitz (2024): "Corporate Loan Spreads and Economic Activity," *The Review of Financial Studies*, 38(2), 507–546.

Wang, Y., T. M. Whited, Y. Wu, and K. Xiao (2022): "Bank Market Power and Monetary Policy Transmission: Evidence from a Structural Estimation," *The Journal of Finance*, 77(4), 2093–2141.