UniCredit Foundation

"A Macroeconomic Model with Heterogeneous Banks"

Rustam Jamilov

7th Econ Job Market Best Paper

Working Paper Series

n. 148 ■ November 2020

Statement of Purpose

The Working Paper series of the UniCredit Foundation is designed to disseminate and to provide a platform for discussion of either work of the UniCredit economists and researchers or outside contributors (such as the UniCredit Foundation scholars and fellows) on topics which are of special interest to the UniCredit Group. To ensure the high quality of their content, the contributions are subjected to an international refereeing process conducted by the Scientific Committee members of the Foundation.

The opinions are strictly those of the authors and do in no way commit the Foundation and UniCredit.

These Working Papers often represent preliminary work. Citation and use of such a paper should take account of its provisional character.

The Working Papers are also available on our website (http://www.unicreditfoundation.org)

A Macroeconomic Model with Heterogeneous Banks

Rustam Jamilov[†]
London Business School

This version: March 15, 2021 First version: March 28, 2020

Abstract

I build a dynamic general equilibrium model with a banking sector that features uninsurable idiosyncratic rate of return shocks, monopolistic credit markets with endogenously variable markups, costly default, and endogenous entry. The model yields empirically realistic distributions of bank assets, net worth, deposits, leverage, markups, relative prices, deposit rates, and default risk. In line with the data, in this environment larger banks are more efficient due to the economies of scale, choose to charge higher markups, and have lower distance to insolvency than smaller banks. The model highlights a *trilemma* for bank regulation: the government cannot simultaneously improve financial competition, efficiency, and stability. Heterogeneous capital requirements enhance stability but reduce efficiency. Deposit insurance schemes stimulate lending and investment but increase equilibrium risk-taking and fragility. Constrained-efficient allocations, which internalize the impact of all bank-level choices on aggregate returns, elevate gross welfare but considerably reduce ex-post stability. I discuss predictions and implications of the framework for the ongoing rise of banking concentration, emergence of fintech credit, the "too-big-to-fail" hazard, implementation of targeted stabilization policies such as bank-level bailouts and liquidity facilities, and intermediary asset pricing.

Keywords: Macro-Finance, Heterogeneous Intermediaries, Bank Market Power

JEL Codes: E44, G20, G21, G28, G32, G38

[†]Department of Economics, London Business School. Web: www.rustamjamilov.com. E-mail: rjamilov@london.edu. I am indebted to Hélène Rey for the invaluable guidance and support. I am grateful to the discussant, Itamar Drechsler, for very helpful comments and suggestions. I thank Michele Andreolli, Fernando Broner, Nuno Coimbra, Dean Corbae, Briana Chang, Ester Faia, Miguel Faria-e-Castro, Andrea Ferrero, Luca Fornaro, Xavier Gabaix, Sigurd Galaasen, Francois Gourio, Francisco Gomes, Veronica Guerrieri, Kyle Dempsey, Tarek Hassan, Zhiguo He, Kilian Huber, Luigi Iovino, Ragnar Juelsrud, Diego Kaenzig, Anil Kashyap, Nobu Kiyotaki, Ralph Koijen, Joseba Martinez, Michael McMahon, Frederic Mishkin, Tommaso Monacelli, Simon Mongey, Plamen Nenov, Tsvetelina Nenova, Elias Papaioannou, Pascal Paul, Franck Portier, Albert Queralto, Morten Ravn, Ricardo Reis, Kenneth Rogoff, Petr Sadlacek, Andrew Scott, Vania Stavrakeva, Vincent Sterk, Kjetil Storesletten, Ludwig Straub, Paolo Surico, Jenny Tang, John Vickers, Annette Vissing-Jorgensen, Randall Wright, Francesco Zanetti, Tong Zhang as well as seminar participants at Oxford, UC Berkeley, University of Wisconsin, New York Fed, Chicago Fed, Boston Fed, Board of Governors of the Federal Reserve, Queen Mary University of London, BI Norwegian Business School, Junior VMACS Conference, AEA 2021, EEA 2020, EFA 2020, ESWM 2020 for helpful feedback. I thank the UniCredit Foundation, the AQR Asset Management Institute, and the Wheeler Institute for Business and Development for financial support. Part of this research was conducted while I was visiting the Research Department of Norges Bank, whose hospitality I gratefully acknowledge. All errors are my own.

1 Introduction

Is there a trade-off between competition, efficiency, and stability in the modern banking system? This paper argues that we should think of these three dimensions through the lenses of a "trilemma": any policy intervention that enhances one of these structural facets necessarily exacerbates one or more of the remaining two. This is a simple and novel generalization of the canonical financial competition-stability debate in a world where banks differ systematically in their profitability, market power, marginal costs, risk profiles, and lending capacity. The trilemma has immediate implications for new trends in banking such as the rise of concentration and emergence of fintech-intermediated credit. It also offers a fresh perspective on classic issues in bank regulation like the "too-big-to-fail" hazard, deposit guarantee schemes, and capital requirements. Furthermore, the trilemma can guide practical implementation of unconventional monetary and fiscal tools such as targeted bailouts and liquidity facilities.

The trilemma arises naturally in a tractable macroeconomic model with a banking sector that is consistent with the following four motivating facts:

Fact 1: *The banking industry is highly concentrated.* Moreover, the industry is becoming more concentrated over time. This is true for the U.S. as well as for the Euro area (Corbae and D'Erasmo, 2020b; Constancio, 2016). As of 2021, the 10 largest banks in the U.S. control roughly 60% of the nationwide loan market.

Fact 2: *There are economies of scale in lending; larger banks are more efficient than smaller banks.* Multiple empirical studies have confirmed presence of either cost- or productivity-driven economies of scale in banking (Wheelock and Wilson, 2012, 2018; Berger and Mester, 1997; Berger and Hannan, 1998). As a bank's balance sheet grows, both marginal and fixed costs begin to shrink relative to assets under management. Economies of scale is also the cornerstone of core principles in canonical banking theory such as delegated monitoring (Diamond, 1984).

Fact 3: *Larger banks charge higher loan markups than smaller banks.* This relatively novel stylized fact has appeared in the works of Corbae and D'Erasmo (2020a) and Pasqualini (2021). Authors apply a variant of the production-function approach that De Loecker et al. (2020) have popularized for the study of market power and find that bank markups are concentrated in the right tail of the size distribution. Elsewhere, Benetton (2021) in a structural model of the UK mortgage market also finds that larger banks charger higher loan markups.

Fact 4: *Bank defaults are costly for the real economy.* The literature on the aftermath of financial crises is vast and some of the notable contributions include Reinhart and Rogoff (2009), Schularick and Taylor (2012), Romer and Romer (2017), and Laeven and Valencia (2018). The literature consensus is that financial crises, especially systemic banking crisis episodes, lead to considerable,

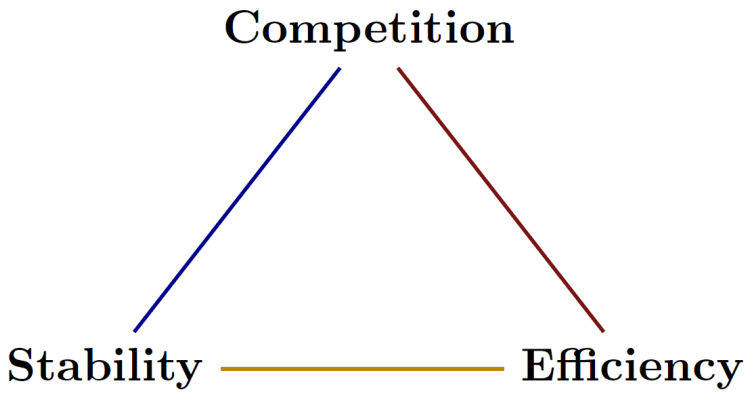
prolonged declines in economic activity, financial intermediation, and consumer welfare.

To formalize these facts into a coherent framework, I build a parsimonious dynamic general equilibrium macroeconomic model with heterogeneous banks. There are four main buildings blocks to this quantitative theory. First, we start with a stripped-down version of the workhorse representative-bank macroeconomic environment of Gertler and Kiyotaki (2010) and Gertler and Karadi (2011) and nest them as a special case. Second, banks engage in monopolistic competition in the credit market with non-CES demand, as in Kimball (1995). This is a tractable way to engineer variable loan markups that increase with relative balance sheet size. This setup nets the Dixit and Stiglitz (1977) (CES) aggregator as a special case and has been applied widely in the literature on monopolistic competition with non-financial firms (Klenow and Willis, 2016; Midrigan et al., 2018). To the best of my knowledge, mine is the first attempt to apply this simple but very effective modelling technique to the case of financial market power. Third, banks face partially uninsurable idiosyncratic rate of return risk in the spirit of Benhabib et al. (2019). This assumption is motivated, among others, by the recent empirical work of Galaasen et al. (2020) who find, using administrative loan-level data from Norway, that idiosyncratic firm shocks survive portfolio aggregation and have a significant impact on bank returns and the aggregate economy. Uninsurable idiosyncratic shocks create an exogenous bank net worth fluctuation problem analogous to the canonical Bewley-Huggett-Aiyagari-Imrohoglu environment (Bewley, 1977; Huggett, 1990; Aiyagari, 1994; Imrohoglu, 1996). Finally, idiosyncratic risk is a source of insolvency risk for banks, who can default on their short-term debt obligations. Default is costly and the cost increases with the size of the bank.

The model works in the following way. Persistent idiosyncratic rate of return shocks drive ex-post heterogeneity in bank net worth (size, for short). Due to scale variance and model non-linearities, the joint distribution of idiosyncratic returns and size is now a state variable: banks with different profitability-size combinations make different balance sheet choices that include risky investments, short-term debt from households, and markups. Each bank also faces its own equilibrium probability of default due to insolvency. When there is no deposit insurance, the cross-section of default risk is priced competitively into the distribution of deposit rates. In general equilibrium, we obtain a collection of stationary right skewed distributions of bank net worth, assets, book and market leverage ratios, markups, relative prices, default probabilities, and deposit rates.

The model produces several cross-sectional predictions that are both intuitive and consistent with the data. First, small-net-worth banks are those who (a) have a poor history of idiosyncratic return realizations (they are "unlucky"), (b) have shorter distance to default, and (c) face higher equilibrium interest rates on short-term debt. All three factors contribute to smaller banks facing higher marginal costs. Stationary marginal cost heterogeneity drives the *economies of scale channel*

Figure 1: **The Banking Industry Trilemma**



Notes: Figure visualizes the competition-efficiency-stability trade-off that arises in the model. Competition stands for the average equilibrium credit markup. Efficiency is the average marginal cost or the inverse of the credit supply elasticity of banks. Stability is the average probability of bank default due to insolvency.

- bigger banks have a higher lending capacity because they are endogenously more cost-efficient. As a result, the credit supply elasticity with respect to net worth shocks increases with the level of initial net worth and decreases with the marginal cost. Second, as mentioned previously, smaller banks face a higher equilibrium probability of default. This is the *financial stability channel*. Third, because of the Kimball aggregator, high-net-worth banks choose to charge higher markups. This is the endogenous *competition channel*. As a result, in the concentrated stationary distribution of net worth we get that bigger banks are more efficient due to economies of scale, charge higher markups, and default less often than smaller banks. Variations in the extensive margin, i.e. the right tail of the distribution, can impact total lending capacity, the average markup, and systemic risk (average probability of default). Banks in our model are "atomistic" (because of monopolistic competition) but still "granular" in the sense that concentration of the distribution matters directly for macroeconomic outcomes (Gabaix, 2011).

We can now discuss the *banking industry trilemma* that has appeared endogenously and fully in general equilibrium. Figure 1 helps to visualize the result. Because idiosyncratic bank return shocks are volatile and persistent, the equilibrium distribution of bank size is right-skewed and concentrated. Relative to a representative-bank counterfactual, this economy is more efficient, more stable, but less competitive. The regulator can potentially improve on the market allocation through a myriad of policy interventions. I consider several viable possibilities. First, regulatory capital requirements that increase with bank leverage can improve financial stability by reducing the aggregate leverage ratio and systemic risk. However, this intervention meddles with the banks precautionary lending motive and their ability (and desire) to grow. As a result, the regulated economy is less efficient as aggregate lending falls and costs rise. Second, introducing full deposit insurance generally has a positive effect on lending and growth but a large negative effect on

stability. Finally, constrained-efficient allocations and optimal heterogeneous bank taxation, which fully internalize the impact of all bank choices (assets, debt, markups) on aggregate prices and returns, improve gross welfare but severely worsen systemic financial stability. When default is sufficiently costly ex-post, net welfare effects could be negative. I conclude that as long as there is a trilateral trade-off between competition, efficiency, and stability of the form that the model generates, the regulator will rarely if ever be able to improve all three dimensions of the banking system simultaneously. It is a trilemma. Bank heterogeneity is critical for this intuition to go through: it is important that bank net worth is positively correlated with distance to default (high stability) and markups (low competition) but negatively correlated with marginal costs (high efficiency).

In the rest of the paper, I apply the framework and the trilemma to several old and new pressing issues in macro-banking. First, the global banking industry is becoming more and more concentrated. My theory predicts that this permanent "granularity" shock will make the banking system more efficient, stable, but less competitive. This conjecture follows immediately from the combination of intensive and extensive margins: bigger banks charge higher markups but are less prone to default and are more efficient. Second, the worldwide share of fintech and bigtech in financial intermediation is growing rapidly (Claessens et al., 2018). My framework predicts that the emergence and rise of fintech credit will lead to economic growth, a significant increase in the number of active banks, but ultimately a decline in financial stability since the economy would be populated with too many small and risky intermediaries which lack the scale to withstand idiosyncratic uncertainty. Third, the "too-big-to-fail" problem is one of the most salient features of modern banking systems in recent decades. A model with rich heterogeneity like ours makes it possible to operationalize the TBTF moral hazard in full general equilibrium. The framework predicts that when a TBTF subsidy is fully priced and internalized by all agents in the economy, all macroeconomic aggregates like lending and production decline while systemic fragility goes up. This finding is consistent with the idea that the TBTF subsidy makes private leverage choices of individual banks strategic complements (Farhi and Tirole, 2017).

The model has two final auxiliary implications that could be useful in their own right. First, on the implementation of various unconventional, bank-level stabilization policies that have become very popular since the 2007-2008 Great Financial Crisis. I find that the impact on aggregate output of targeted policies depends on the policy type and on the region of the banking distribution that is being affected. Policies that shift relative prices or marginal costs are generally more effective when applied only to small banks. These policies include targeted lending and liquidity facilities. This is because relaxation of marginal cost pressures benefits smaller intermediaries by more, as large banks' funding costs are relatively low. On the other hand, aggregate efficiency gains from unanticipated targeted equity injections ("bailouts") generally *increase* with the size of the

impacted intermediary. This is because, in equilibrium, credit supply elasticities are greater for high-net worth banks who face low marginal costs. This is due to the economies of scale channel.

Second and finally, the model offers a fresh perspective on intermediary asset pricing with heterogeneity. My framework can generate a sizable unconditional risk premium and *countercyclical* risk premia. The correct risk premium in the model is the Mehra et al. (2011) "intermediation spread" - the difference between the return on risky investment by banks and the theoretical risk-free rate. The spread is zero without any financial frictions. The spread is positive in equilibrium thanks to two channels. First, the occasionally binding leverage constraint that banks face is the source of a *liquidity premium*. The second one is the *default risk premium*. Liquidity and default risk premia of the representative bank, with no heterogeneity, add up to roughly 2%. Introducing uninsurable idiosyncratic shocks makes both liquidity and default risk premia heterogeneous and concentrated in the left tail of the distribution. A significant equilibrium fraction of small and risky intermediaries is responsible for increasing the total risk premium by another 2%. Finally, the whole distribution of bank net worth is countercyclical. Therefore, the share of banks with high liquidity and default risk premia increases precisely in high marginal utility states. This makes the aggregate risk premium countercyclical endogenously and *without* the assumption that idiosyncratic risk is exogenously greater in recessions.

Literature. This paper contributes to several literature strands.

First, I build on a long literature that studies the tradeoffs between financial competition, stability, and growth (efficiency). One view is that competition reduces franchise values of the banks and induces more risky behavior. (Keeley, 1990; Hellman et al., 2000; Repullo, 2004; Beck et al., 2006). There is also an alternative view that riskiness of loans correlates with the level of the interest rate. As a result, greater competition may reduce default risk (Boyd and Nicolo, 2005; Martinez-Miera and Repullo, 2010). My contribution is to synthesize the tradeoffs, both in a positive and normative sense, in a dynamic general equilibrium environment with endogenously variable bank markups and costly bank default.[1]

Second, several studies have emphasized the role of financial heterogeneity and rely, like my model does, on some form of idiosyncratic risk and ex-post heterogeneity. Such papers include Corbae and D'Erasmo (2020a), Bianchi and Bigio (2020), Rios Rull et al. (2020). Rios Rull et al. (2020) study countercyclical capital buffers in a partial-equilibrium setting with idiosyncratic bank

[1]Articles that look at imperfect financial competition in equilibrium setups include Christiano and Ikeda (2013), Nguyen (2015), Stavrakeva (2019), Capelle (2019), Davydiuk (2019). A growing literature also looks at imperfect competition in the market for bank deposits and liabilities in general, a channel that we abstract from in this paper (Drechsler et al., 2017; Egan et al., 2017). Corbae and Levine (2018) review the state of the literature on financial competition in their 2018 Jackson Hole Symposium address. Their work stresses the theoretical interactions between competition, financial fragility, and monetary policy.

default risk. Bianchi and Bigio (2020) study competitive banks' liquidity management problem in a model with idiosyncratic deposit withdrawal shocks. Corbae and D'Erasmo (2020a) study oligopolistically competitive banks that are subject to idiosyncratic shocks on the liability side of the balance sheet. My main contributions are twofold. First, I study market power and heterogeneity stemming from the asset side of the banks balance sheet with a highly flexible monopolistic financial competition setup that can accommodate both constant (CES) and variable (Kimball) markups. Second, I explore normative implications in a realistic but complex environment with incomplete markets, variable bank markups, and default risk.

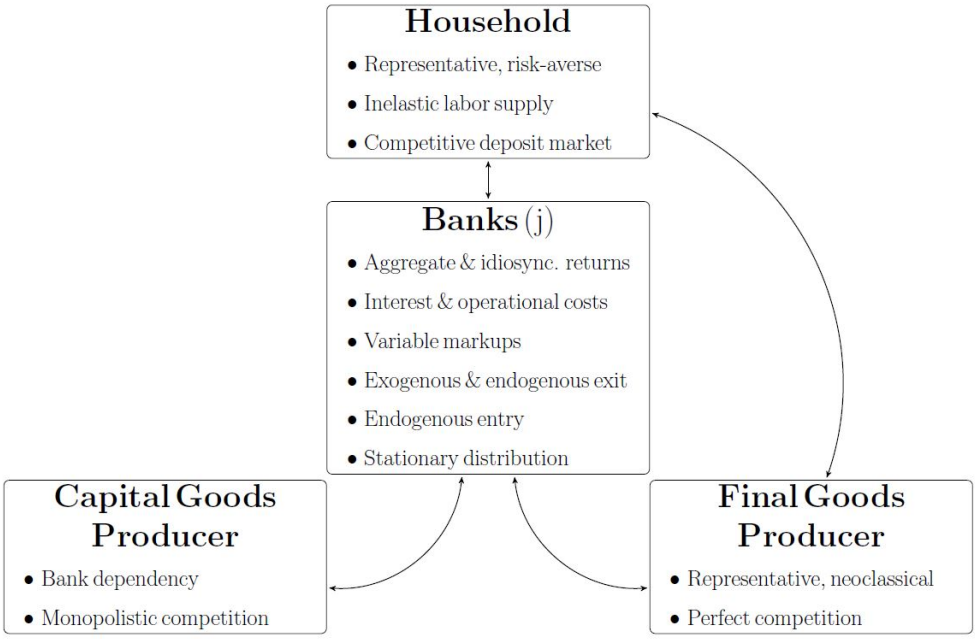
Third, my model is related to the literature that introduces *ex-ante* heterogeneity among financial intermediaries. Coimbra and Rey (2019) develop a general equilibrium model with ex-ante heterogeneity in intermediary value-at-risk constraints and endogenous financial stability. Their model features, like ours, dynamic intensive and extensive margins of bank risk-taking. My approach differs from theirs in two substantial ways. First, in my model market incompleteness and uninsured idiosyncratic return risk achieve persistent *ex-post* heterogeneity of bank returns and balance sheet characteristics. Second, my model departs from the assumption of perfect competition in bank lending. This channel delivers rich ex-post variation in markups and relative prices.[2]

Finally, this paper contributes to a long-running literature that introduces credit frictions and financial intermediaries into general equilibrium macroeconomic models. I build on the workhorse macro-banking setup of Gertler and Kiyotaki (2010) and Gertler and Karadi (2011), whom my model nests as special cases. A very incomplete list of salient equilibrium models with a financial sector includes Gromb and Vayanos (2002), Brunnermeier and Pedersen (2009), Adrian and Shin (2010, 2014), Jermann and Quadrini (2013), Brunnermeier and Sannikov (2014), He and Krishnamurthy (2013), Nuno and Thomas (2016), Gertler et al. (2016, 2020), Lee et al. (2020), Bigio and Sannikov (2021) etc. These frameworks largely abstract from distributional considerations in the financial sector and work with a representative financial intermediary/entrepreneur/arbitrageur whose relative scale generally cannot be pinned down.

Outline. The rest of the paper is structured as follows. Section 2 lays out the model. Section 3 discusses how we take the model to the data. Section 4 presents alternative bank regulation schemes and tools. Section 5 is on various practical and empirical applications. Finally, Section 6 concludes.

[2]Other papers that develop models with ex-ante financial heterogeneity include Korinek and Nowak (2016), Boissay et al. (2016), Begenau and Landvoigt (2020), Begenau et al. (2020), Goldstein et al. (2020), Dempsey (2020).

Figure 2: **Model Overview**



2 Model

In this section, I present the model, discuss its key building blocks, and analyze equilibrium properties.

2.1 Overview

Figure 2 visualizes the structure of the model. Time is discrete and infinite. The economy consists of a representative household, a continuum of financial intermediaries that are ex-ante identical but ex-post heterogeneous, a representative final goods producer, and a capital goods producer. The representative household is risk-averse, supplies labor inelastically to the final goods firm in exchange for a competitive wage, and saves intertemporally through one-period bank deposits. The deposit market is perfectly competitive and there is no deposit insurance in the baseline economy; we will introduce it in Section 4. Banks accumulate own net worth, acquire deposits from households to whom they pay the equilibrium deposit rate the following period, cover non-interest expenses that are non-linear in assets under management, and perform two investment activities. First, banks invest into claims on zero-profit capital goods firms who produce aggregate capital. Competition is monopolistic and demand is non-CES (Kimball).[3] After the capital stock is

[3]Why don't we set up an oligopolistic credit market? There are two reasons. First, although the number of active banking franchises in the U.S. and Europe is dwindling, there are still more than 5,000 active commercial banks in the U.S. at the time of writing. A monopolistic competition structure feels much more appropriate given the number

produced and priced, banks immediately lend it competitively to the final goods producer in return for realized returns on capital the following period.[4] In addition to the systematic return, each bank receives a bank-specific idiosyncratic return draw. These shocks are persistent and not perfectly insurable. Along the extensive margin, banks exit due to exogenous and endogenous reasons. Upon exogenous exit, banks pass on all remaining net worth to households, which motivates our constant dividend payout rule. Endogenous default is due to fundamental insolvency. Finally, entry is exogenous in the baseline economy. Entry and the mass of active banks are endogenized in Section 5.[5]

An important advantage of the model is that the two essential building blocks - monopolistic financial competition and uninsurable idiosyncratic bank return shocks - can be shut down without affecting the other. In other words, we can analyze an economy with heterogeneity but perfect competition, monopolistic competition but homogeneity, or any re-calibrated combination in between.

2.2 Technology

Final Good Production The final good is produced from aggregate capital and labor using a Cobb-Douglas technology:

$$Y_t = AK_t^{\alpha} L_t^{1-\alpha} \tag{1}$$

where $0 < \alpha < 1$. Wages (W_t) and returns to capital (R_t^k) are competitive and follow directly from the firm's optimization problem:

$$R_{t+1}^k = \frac{A\alpha K_{t+1}^{\alpha-1}}{P_t} \qquad W_t = (1-\alpha)AK_t^{\alpha} \tag{2}$$

where A is aggregate productivity and P_t is the price of capital. Capital depreciates fully every period after it's used in production.

Capital Good Production A representative, perfectly competitive capital producing firm begins the period with no equity and issues claims to banks in return for the aggregate capital bundle. The

of agents in the market. In addition, it is more reasonable that individual banks do *not* internalize the impact of their private choices on aggregate outcomes. In other words, no matter how big or "granular" banks can become, they are still atomistic. This is what monopolistic competition grants us. Second, tractability. There is practically nothing that an oligopoly model can give us that the monopolistic competition block with variable markups cannot. A flexible two-parameter departure from perfect competition cannot be taken for granted in an environment with uninsurable idiosyncratic shocks like ours.

[4]Ownership of the capital stock is in pro rata terms.

[5]In Section B.3 we show how the baseline economy can be extended to feature two sectors that are heterogeneous in the degree of financial competition.

firm makes zero profits.

The capital good K_t is produced from a bundle of financial varieties $k_t(j)$ for $j \in [0, 1]$. Financial varieties are intermediated by the banking sector and assembled with the *Kimball* aggregator:

$$K_t = \int_0^1 \Upsilon\left(\frac{k_t(j)}{K_t}\right) dj \tag{3}$$

where the function $\Upsilon(x)$ is increasing, concave, and satisfies $\Upsilon(1) = 1$. It can be shown that the Dixit-Stiglitz aggregator is a special case with $\Upsilon(x) = x^{\frac{\theta-1}{\theta}}$, where $\theta > 1$ is the constant elasticity of substitution.

The maximization problem of the capital goods firm is:

$$\max_{k_t(j)} \left[P_t K_t - \int_0^1 p_t(j) k_t(j) dj\right]$$

subject to technology 3. This yields a demand function for bank funds:

$$p_t(j) = \Upsilon'\left(\frac{k_t(j)}{K_t}\right) Z_t \tag{4}$$

where

$$Z_t := \left(\int_0^1 \Upsilon'\left(\frac{k_t(j)}{K_t}\right) \frac{k_t(j)}{K_t} dj\right)^{-1} \tag{5}$$

is the *Kimball demand index*. In the Dixit-Stiglitz special case, $Z_t = \frac{\theta}{\theta-1}$, and (4) reduces to $p_t(j) = \left(\frac{k_t(j)}{K_t}\right)^{\frac{-1}{\theta}} P_t$.

Discrete Choice Microfoundation It is possible to theoretically underpin the monopolistic credit demand system above using discrete choice theory where each borrower chooses both the size of the loan and the bank/variety to borrow from (McFadden, 1984). The approach generalizes the case of a representative capital goods producer to a large number of borrowers that are heterogeneous in their preferences for individual banks. In other words, there are firm-bank fixed-effect shocks. These shocks are cross-sectionally correlated and the degree of correlation maps into the constant elasticity of substitution θ. Appendix B provides a detailed guide for the analytically more convenient case of $\epsilon = 0$.

Market power at the level of a bank can now be viewed as being isomorphic to consumer (firms, in this case) preferences for financial services that are not perfect substitutes across banks. Even if a particular bank offers higher loan rates (lower prices on claims), it can still remain in business if borrower-bank-specific preference shocks are sufficiently diverse. The problem of heterogeneous

firms is static. In our dynamic setting, as long as the distribution of preferences is not dynamic or aggregate state-dependent, the identical problem would yield the same solution every period. We therefore proceed working with this representative-firm approximation of the more sophisticated heterogeneous-firms environment that is understood to be operating in the background.

2.3 Banks

The credit demand system in (3)-(5) is taken as given by every bank. Intermediaries start the period with initial net worth $n \in \mathbf{N} \subset \mathbf{R}_+$ and must choose assets k(j), deposits d(j), and price of claims p(j) while respecting the balance sheet constraint:

$$d_t(j) + n_t(j) = p_t(j)k_t(j) \tag{6}$$

Every bank faces non-interest expenses $\frac{1}{\zeta_1}k_t(j)^{\zeta_2}$ where parameter ζ_2 can help govern the degree of non-linearity and scale-variance. When choosing the size of the balance sheet, banks can borrow deposits d(j) from the household, subject to the bank-specific interest rate $\bar{R}_t(j)$ that will be determined in general equilibrium.

At the end of each period, every bank earns realized returns on claims on the final goods firm. Each bank earns a portfolio return $R_t^T(j)$ that comprises the return on aggregate capital R_t^k, which is common to all j, and an idiosyncratic component $\xi_t(j)$ which is specific to each bank:

$$R_t^T(j) = \kappa\xi_t(j) + (1-\kappa)R_t^k \tag{7}$$

Where $0 < \kappa < 1$ is a parameter that governs the ability to hedge idiosyncratic shocks. We discuss a possible microfoundation for the $R^T(j)$ formulation in Appendix B.2. The idiosyncratic return, $\xi \in \Xi$, follows an AR(1) process:

$$\xi_t(j) = (1-\rho_\xi)\mu_\xi + \rho_\xi\xi_{t-1}(j) + \sigma_\xi\epsilon_t(j) \tag{8}$$

Analogously, let the finite state Markov representation of (8) be $\mathbf{G}(\xi_{t+1}, \xi_t)$. We can now state the law of motion of bank-level net worth:

$$n_{t+1}(j) = R_{t+1}^T(j)p_t(j)k_t(j) - \bar{R}_t(j)d_t(j) - \frac{1}{\zeta_1}k_t(j)^{\zeta_2} \tag{9}$$

Following Gertler and Karadi (2011) and Gertler and Kiyotaki (2010), there is a moral hazard problem in the deposit market. The bank has an incentive to divert franchise assets with the ability to divert no more than a fraction λ of the total value of revenues p(j)k(j). If deciding to divert, the banker always escapes but the franchise enters bankruptcy the following period. The banker is

indifferent between operating honestly and diverting when whatever he is able to finance exactly equals the value of the franchise. This yields the following incentive constraint that puts a limit on bank leverage.

$$\lambda p_t(j)k_t(j) \leq V_t(j) \tag{10}$$

where $V_t(j)$ is the franchise value of the intermediary, to be defined below. Each bank in the distribution can default with an endogenous probability $\upsilon(j)$. Default risk is due to fundamental insolvency, i.e. when net worth at normal market prices is non-positive.:

$$\upsilon_t(j) = \Pr\Big(n_{t+1}(j) \leq 0\Big) \tag{11}$$

Conditional on insolvency, the household recovers a fraction of promised payments $x_t(j)$, an object that we define later. Because at normal market prices the recovery rate $x_t(j)$ is increasing in bank size, insolvency risk is concentrated in the *left* tail of the stationary bank net worth distribution.

Let $\eta(n, \xi)$ be a probability measure, defined on the Borel algebra B that is generated by open subsets of the product space $\mathbf{B} = \mathbf{N} \times \Phi$, corresponding to the distribution of incumbent banks with net worth n and idiosyncratic return realizations ξ. The law of motion for the distribution is:

$$\eta_{t+1}(n_{t+1}, \xi_{t+1}) = \Phi(\eta_t) \tag{12}$$

We define Φ in detail below.

Dynamic Problem of the Incumbent Banker The following summarizes the dynamic problem of the incumbent. We adopt recursive notation because the solution does not depend on a specific bank j but on the relevant state variables only. Define $\mathbf{s} = \{n, \xi\}$ as the bank's idiosyncratic state vector. There is no aggregate risk. The bank maximizes its franchise value which is defined as the discounted stream of future flows of net worth. With an exogenous probability σ the incumbent may exit involuntarily, in which case its net worth gets transferred lump sum to the household. The banker discounts the future by adopting and augmenting the household's stochastic discount factor Λ, which is determined in equilibrium and defined when we discuss the household problem. Each banker takes as given aggregate quantities $\{K, D, N\}$, prices $\{P, R^k\}$, the cross-sectional distribution $\{\eta\}$, bank-specific deposit rates $\{\bar{R}\}$ and portfolio returns $\{R^T\}$, and the law of motion of the distribution Φ. Each bank solves:

$$V(\mathbf{s}) = \max_{\{k,p,d\}} \left\{ \mathbb{E}_{\mathbf{s'|s}} \left[\Lambda' \Big((1-\sigma)n' + \sigma V(\mathbf{s}') \Big) \right] \right\} \tag{13}$$

s.t. conditions 3-12.

We can simplify the problem above considerably by reformulating it into a one-argument problem. Each bank now chooses the leverage ratio $\phi = \frac{pk}{n}$ by maximizing:

$$\max_{\phi} \left[\mu_a \phi + \nu_a\right] \tag{14}$$

subject to the same constraints as before and where $\mu_a = (1-\upsilon)\tilde{\Lambda}' \left[R^{T'} - \bar{R}\right]$ is the excess return on risky investments, $\nu_a = (1-\upsilon)\tilde{\Lambda}' \left[\bar{R} - \frac{\frac{1}{\zeta_1} k_t(j)^{\zeta_2}}{n}\right]$ is the cost of liabilities. In both instances, $\tilde{\Lambda}' = \Lambda\left(1 - \sigma + \sigma V(\mathbf{s}')\right)$ is the augmented household marginal rate of substitution.

Appendix A.1 shows that the solution to the above problem, while taking all aggregate quantities and equilibrium prices as given, yields the following relative price rule:

Proposition 1 (Markups and Marginal Costs Decomposition)**.**

$$\frac{p(j)}{P} = \mu(x) \frac{k(j)^{\zeta_2 - 1}}{R^T(j) - \bar{R}(j)} \tag{15}$$

where $\mu(x)$ is a markup function, which potentially depends on relative size $x = \frac{k(j)}{K}$, and $\frac{k(j)^{\zeta_2-1}}{R^T(j)-\bar{R}(j)}$ the endogenous marginal cost. In the two paragraphs that follow, we zoom in on these two sources of bank heterogeneity in the model: markups and marginal costs.

2.4 Variable Markups

For the baseline case with endogenously variable bank markups, I use the Klenow and Willis (2016) specification for $\Upsilon(x)$:

$$\Upsilon(x) = 1 + (\theta - 1) \exp \frac{1}{\epsilon} \epsilon^{\frac{\theta}{\epsilon} - 1} \left[\Gamma\left(\frac{\theta}{\epsilon}, \frac{1}{\epsilon}\right) - \Gamma\left(\frac{\theta}{\epsilon}, \frac{x^{\frac{\epsilon}{\theta}}}{\epsilon}\right)\right] \tag{16}$$

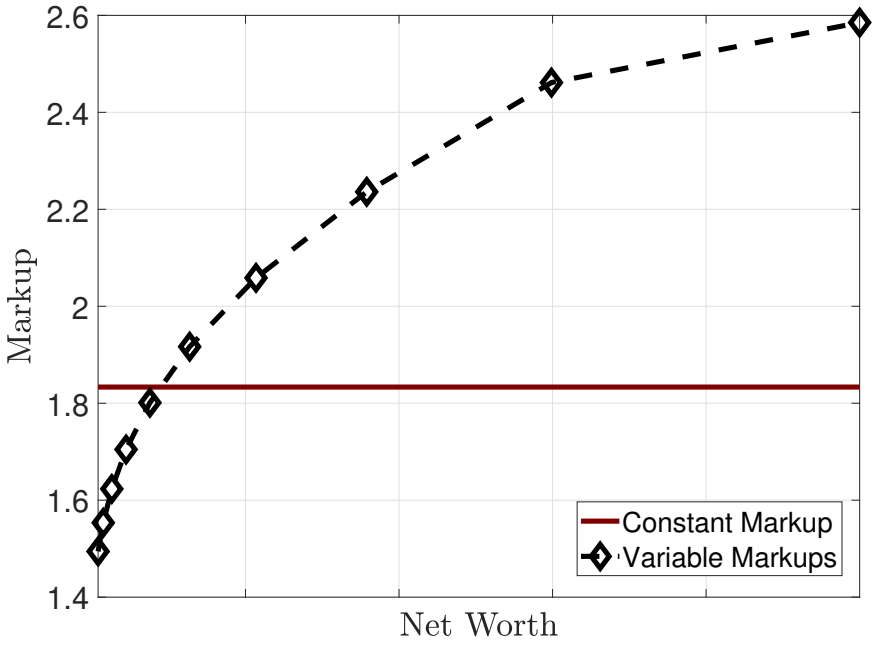
where $\epsilon \geq 0$ is a parameter that governs variation in the *superelasticity* and $\Gamma(s, q)$ is the upper-incomplete Gamma function:

$$\Gamma(s, q) := \int_q^{\infty} t^{s-1} \exp^{-t} dt \tag{17}$$

The CES aggregator is a special case of (16) when $\epsilon = 0$. With the Klenow-Willis specification, we have:

$$\Upsilon'(x) = \frac{\theta - 1}{\theta} \left(\exp \frac{1 - x^{\frac{\epsilon}{\theta}}}{\epsilon}\right) \tag{18}$$

Figure 3: **Bank Markups**



Notes: Absolute bank markups with the Kimball ("Variable Markups"') and CES ("Constant Markup") aggregators.

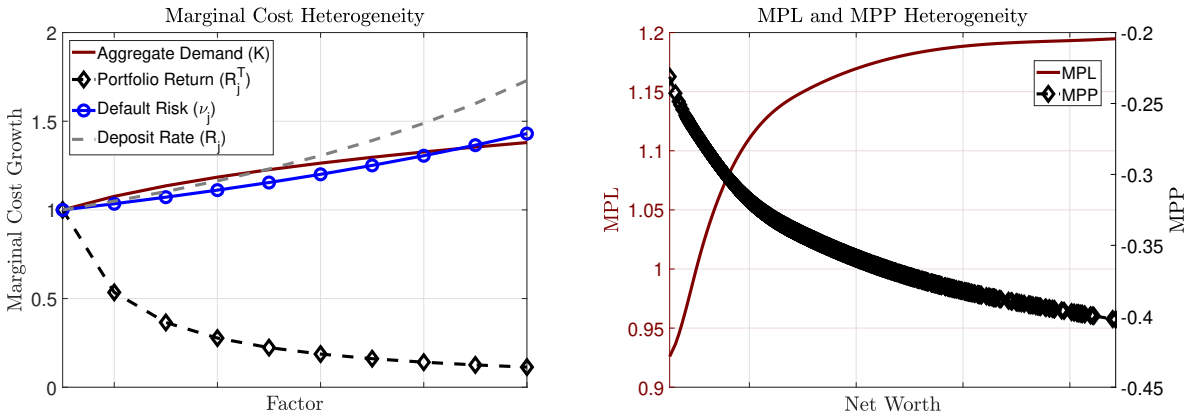
The size-dependent elasticity is thus $\theta \mathrm{x}^{-\frac{\epsilon}{\theta}}$. It can be seen clearly that the elasticity declines with relative size. This, in turn, implies the following markup function:

$$\mu(\mathrm{x}) = \frac{\theta \mathrm{x}^{-\frac{\epsilon}{\theta}}}{\theta \mathrm{x}^{-\frac{\epsilon}{\theta}} - 1} \tag{19}$$

As long as $\epsilon > 0$, banks with a higher relative quantity of assets on their books ($\mathrm{x} = \frac{\mathrm{k}}{\mathrm{K}}$) will face a lower elasticity of substitution. This, in turn, induces larger banks to choose higher $\mu(\mathrm{x})$. When $\epsilon = 0$, the credit markup is constant and equals the usual $\mu = \frac{\theta}{\theta - 1}$. Calibration of the superelasticity can be achieved in a simple way by varying $\frac{\epsilon}{\theta}$. When taking the model to the data, we will use the empirical cross-section of bank markups to deduce the two parameters.

Figure 3 illustrates the differences between Kimball-Klenow-Willis and Dixit-Stiglitz aggregators. Increasing ϵ makes the demand curve less "convex", everything else equal. Larger banks are in the area of relative satiation. Because they face lower substitution elasticities, they choose to charge higher markups since further reduction of relative prices does not induce the same desirable quantity effect.

Figure 4: **Marginal Costs and Economies of Scale**



Notes: Left picture shows how bank marginal costs depend on aggregate demand, bank-level portfolio return draw, bank-level probability of default, and bank-level equilibrium deposit rate. Right picture plots marginal propensities to lend and marginal propensities to price as a function of bank net worth.

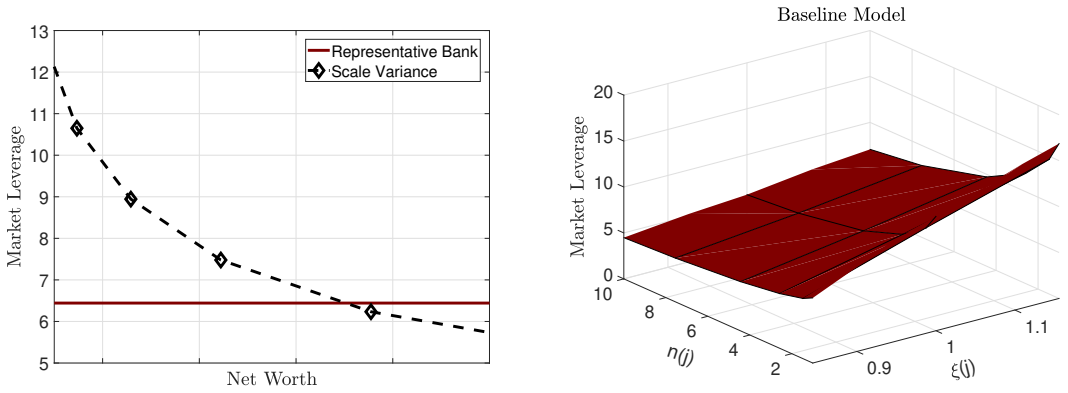
2.5 Marginal Costs and Economies of Scale

Cross-sectional bank heterogeneity in the model also runs through marginal costs. The marginal cost is a complex non-linear function of four key objects: total portfolio return $R^T(j)$, interest rate on deposits $\bar{R}(j)$, the scale effect in K, and the probability of default $\nu(j)$ (which affects marginal costs indirectly, through $\bar{R}(j)$). Note that dependency on aggregate demand K is only possible when non-interest expenses are not linear with respect to k(j), a condition we return to in the next section where we discuss scale variance. The effect of each of the four determinants on the marginal cost is summarized on the left panel of Figure 4. First, we observe that bank-level marginal costs are increasing in aggregate demand. Greater demand for bank finances puts upward pressure on the cost of funds. Second, marginal costs increase in both default risk and interest rates on deposits. Because there is no deposit insurance in the baseline economy, the two are intricately linked and have the same effect on the total marginal cost. Finally, the marginal cost is decreasing in the portfolio return $R^T(j)$, which acts as a profitability shifter and is a source of all ex-post heterogeneity in balance sheet prices and quantities.

Marginal cost heterogeneity gives rise to economies of scale. In order to illustrate the mechanism in the cleanest possible way, we define two new objects: the marginal propensity to lend (MPL) and the marginal propensity to price (MPP). At the level of a bank, MPL(j) is constructed as the elasticity of assets k(j) to changes in net worth n(j). MPP(j) is defined analogously to the MPL as the elasticity of bank-level relative prices p(j) with respect to shocks to bank net worth n(j):

$$\text{MPL} = \int_{\mathbf{B}} \frac{\partial \text{k(j)}}{\partial \text{n(j)}} \eta(\text{dn}, \text{d}\xi) \quad \text{MPP} = \int_{\mathbf{B}} \frac{\partial \text{p(j)}}{\partial \text{n(j)}} \eta(\text{dn}, \text{d}\xi) \tag{20}$$

Figure 5: **Bank Scale Variance**



Notes: Left picture shows how bank leverage depends on net worth in two regimes. "Representative bank" is the case without idiosyncratic shocks and scale variance ($\zeta_2 = 1$). "Scale variance" is the case with scale variance ($\zeta_2 > 1$) and no idiosyncratic shocks. Right picture shows how bank leverage depends on net worth in the baseline economy with both scale variance and idiosyncratic shocks.

The right panel of Figure 4 visualizes the MPL and MPP objects as functions of net worth. We see that MPP(j) and MPL(j) are inversely related, which is due to the Kimball demand function and negative correlation of assets and relative prices. In equilibrium, low-net-worth banks are those that (a) have a poor history of idiosyncratic return realizations, (b) have shorter distance to default, and (c) face higher equilibrium deposit rates. All three factors contribute to smaller banks facing higher marginal costs and being less efficient. High marginal costs, in turn, feed into lower efficiency and lending capacity, which is summarized with a higher (lower) MPL (MPP). This is the economies of scale channel. For contrast, in the representative-bank counterfactual, the MPL and MPP distributions are flat and correspond to the corresponding objects of the median intermediary. Depending on the extensive margin and the relative shares of very large and very small banks in the distribution, heterogeneity becomes important for the transmission of net worth shocks to aggregate lending and investment.

2.6 Scale Variance

We now demonstrate how the baseline economy features scale variance and nests the representative-bank special case. We visualize the mechanism graphically on figure 5. We analyze the optimal choice of bank market leverage $\frac{pk}{n}$ in three different situations. First, we start with the representative-bank case with complete markets (σ_ξ=0 and κ=0), and linear non-interest expenses ($\zeta_2 = 0$). As can be seen from the figure, linearity and complete markets make the leverage ratio one-dimensional and independent of the state of initial net worth.

Second, the downward-sloping line on the left panel of Figure 5 plots optimal leverage for an

extension that allows for scale variance ($\zeta_2 > 1$). Notice how leverage is now decreasing in net worth. Because probability of insolvency in the cross section mirrors market leverage, smaller banks are thus more risky and default more often. Finally, in the right panel of the Figure, we relax the assumption of market completeness. This step introduces ex-post heterogeneity in returns. Moreover, because we continue to retain scale variance, the optimal leverage ratio now depends on two states: $\xi(j)$ and n(j): low-n(j), high-ξ(j) banks choose the highest leverage in the economy.

An important feature of this class of models with financial intermediaries is linearity with respect to net worth. This assumption normally allows the model to be aggregated explicitly. I can formalize the departure from homogeneity by formally proving that the value function of the bank in our model is *not* linear in net worth. In this case, one must track the two-dimensional state of net worth and idiosyncratic risk in addition to aggregate factors such as the aggregate capital stock, because bank-specific characteristics matter for the choice of $\{k(j), p(j), d(j)\}$. This result is in direct contrast to the standard proofs in Gertler and Kiyotaki (2010) and Bocola (2016), among many others. Proposition 2 formalizes the intuition.

Proposition 2 (Bank Scale Variance)**.** *The solution to the incumbent banker's problem, conditional on initial net worth* n(j) *and idiosyncratic return* ξ(j)*, is*

$$V\Big(n(j), \xi(j)\Big) = \vartheta\Big(n(j), \xi(j)\Big) n(j)$$

where the marginal value of net worth is:

$$\vartheta\Big(n(j), \xi(j)\Big) = \frac{\Big(1 - \upsilon(j)\Big)\mathbb{E}\left(\Lambda'\Big[1 - \sigma + \sigma\vartheta\Big(n'(j), \xi'(j)\Big)\Big]\left(\bar{R}(j) - \frac{\frac{1}{\zeta_1}k(j)^{\zeta_2}}{n(j)}\right)\right)}{1 - \varphi\Big(n(j), \xi(j)\Big)}$$

and the multiplier on the moral hazard leverage constraint is

$$\varphi\Big(n(j), \xi(j)\Big) = \max\left[1 - \frac{\Big(1 - \upsilon(j)\Big)\mathbb{E}\left(\Lambda'\Big[1 - \sigma + \sigma\vartheta\Big(n'(j), \xi'(j)\Big)\Big]\left(\bar{R}(j) - \frac{\frac{1}{\zeta_1}k(j)^{\zeta_2}}{n(j)}\right)\right)}{\lambda\phi(j)}, 0\right]$$

Proof: Appendix A.2.

The above proposition shows that the value function is not linear in net worth. This can be seen from the explicit dependency of the marginal value of net worth ϑ on both ξ(j) and n(j). The former is guaranteed by $\kappa > 0$ and $\sigma_\xi > 0$, i.e. idiosyncratic return risk and market incompleteness. The

latter is driven by non-linearity of non-interest expenses in assets under management ($\zeta_2 \neq 1$). As a result, explicit aggregation in the banking sector is not possible as the linearity condition is not satisfied. Financial intermediaries are ex-post heterogeneous in terms of returns, which feeds into all other balance sheet and income statement characteristics.

2.7 Entry and Exit

In the baseline version of the model, entry is exogenous. We relax this assumption and introduce endogenous entry in Section 5. The incumbent intermediary is subject to two sources of exit risk: involuntary homogenous exit rate σ and the endogenous probability of default $\nu(j)$, which is bank-specific. Default is due to fundamental insolvency, which occurs when n(j) is drawn down to 0. Intermediary default is costly and results in ex-post efficiency losses that are measured in units of output. Default costs are potentially bank size-dependent. If a bank exits, the exiting bank's market will never be taken over by any of the incumbents.[6]

2.8 Cross-Sectional Distribution of Banks

Define E_t as the mass of banks that exit the economy due to endogenous default. Recall that (1-σ) is the fraction of banks that draw an exogenous exit shock. Exiting banks are replaced such that the mass of active intermediaries is constant. The distribution of banks in the economy thus evolves according to:

$$\eta'(n', \xi') = \left(\sigma - E'\right) \sum_{\xi} G(\xi', \xi) \int \mathbb{1}_{\{(n,\xi)|K(n,\xi)\in \mathbf{B}\}} \eta(dn, d\xi) \tag{21}$$

Where $\mathbb{1}$ is the indicator function that takes the value of unity when the argument {.} is true and zero otherwise. Recall that $G(x', x)$ is the Markov chain for ξ of the incumbents.

2.9 Households

The representative household is tasked with choosing the supply of deposits to each bank $b_t(j)$ and consumption C_t, subject to the standard balance sheet constraint.

$$\max_{C_t, b_t(j)} \left[\mathbb{E}_t \sum_{t=1}^{\infty} \beta^t u\left(C_t\right) \right] \quad \text{s.t.}$$

$$C_t + \int_0^1 b_t(j)dj \leq W_t + \int_0^1 \bar{R}_t(j) b_{t-1}(j)dj + \pi_t$$

[6] A secondary market for bank mergers and acquisitions is not permitted. Bank runs (liquidity crises) are ruled out.

Where π are any lump sum transfers or taxes. First order conditions for deposits yield the following equation:

$$\bar{R}_t(j) = \frac{1 - \upsilon_t(j) x_t(j) \mathbb{E}\left(R^T_{t+1}(j) \Lambda_{t+1}\right)}{\left(1 - \upsilon_t(j)\right) \mathbb{E}\left(\Lambda_{t+1}\right)} \tag{22}$$

Where $\Lambda_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)}$ is the stochastic discount factor. Deposits are risky because of possible bank default and absence of deposit insurance schemes. The consumer acknowledges default risk and demands a menu of deposit rates, which depend on the deposit recovery rate $x_t(j)$:

$$x_t(j) = \min\left[\frac{\phi_t(j)}{\phi_t(j) - 1}, 1\right]$$

With ϕ the market leverage ratio, defined as before.

2.10 Stationary Industry Equilibrium

Credit market clearing requires:

$$K = \int_{\mathbf{B}} \left(k(n, \xi)\right) \eta(dn, d\xi) \tag{23}$$

Similarly, clearing the deposit market requires:

$$\int_0^1 b(j) dj = \int_{\mathbf{B}} \left(d(n, \xi)\right) \eta(dn, d\xi) \tag{24}$$

The goods market requires the final good to be used either for household consumption or firm investment. The latter includes investment demand that is intermediated both by the incumbent and the entering bankers:

$$Y = C + I$$

We consider equilibria without aggregate uncertainty such that all aggregate quantities, prices, and measures are time-invariant. A *Stationary Industry Equilibrium* is defined as a set of functions that include the value function of the banker $V(\mathbf{s})$, optimal policies for bank capital investment $k(\mathbf{s})$ and deposit demand $d(\mathbf{s})$, household consumption C and deposit supply b(j), competitive wage W and capital R^k pricing functions, the aggregate price of capital P, a marginal utility process Λ, and the menu of market-clearing deposit rates $\bar{R}(\mathbf{s})$ such that:

1. The household's choices $\{C, b(j)\}$ are optimal conditional on $\{W, \bar{R}(j)\}$

2. The banks choices $\{k, p, d, \mu\}$ are optimal conditional on $\{\Lambda, K, P, \bar{R}(\mathbf{s}), \eta\}$

3. Returns on factors of production are: $R^k = \frac{\alpha A K^{\alpha-1}}{P}$, $W = (1-\alpha)AK^{\alpha}$

4. Aggregate quantities $\{K, D, N\}$ are consistent with the cross-sectional distribution and the monopolistic credit demand system

5. The credit market clears as in (23). The deposit market clears as in (24)

6. The cross-sectional distribution evolves according to (21) and is consistent with bank-level demand functions

2.11 Numerical Algorithm

The numerical algorithm that I use to solve the model is described in detail in Appendix E.

3 Taking the Model to the Data

In this section I discuss the parameterization strategy, moments that the model manages or fails to match, and some key cross-sectional patterns in the banking sector.

3.1 Parameterization

All chosen parameters are shown in Table 1. The model period is one quarter. We begin by describing standard macro parameters. We set the share of aggregate capital in production α to 0.36. The discount factor β is chosen to target a steady-state annual risk-free rate of 2.60%. We assume log-utility in consumption.

For parameters in the banking block, we set the exogenous survival probability to $\sigma = 0.9$, which is broadly consistent with the trend exit rate of banks in the U.S. According to the FDIC, there were roughly 11000 commercial banks in the U.S. in 1980. This number has dropped to 5000 by 2019. This implies an average annual exit rate of 3% and a life expectancy of a banker of about 8.25 years. In the model, the expectancy is 10 years. The fraction of divertible assets $\lambda = 0.1$ targets a steady state bank leverage ratio of roughly 7. Endowment of new entrants is set to 30% of average net worth N, which helps to achieve an empirically realistic average entry rate of 5% whenever entry is endogenous. Parameters that govern non-interest expenses (ζ_1, ζ_2) are chosen to be consistent with empirical evidence on increasing returns to scale in banking while allowing the banking problem to remain concave (Wheelock and Wilson, 2018).[7]

[7]Our results do not rely on whether these costs are concave or convex, although convexity is much more computationally convenient. The knife-edge case of $\zeta_2 = 1$ was discussed in Section 2.6.

Table 1: **Parameter Values**

Parameter	Description	Value
	Macro	
α	Share of capital in production	0.36
β	Discount factor	0.996
σ_h	Risk aversion	1
	Banking	
σ	Dividend payout ratio	0.9
ω	Share of divertible assets	0.1
φ	New banker endowment	30% of N
$\frac{1}{\zeta_1}$	Monitoring cost linear	0.01
ζ_2	Monitoring cost quadratic	1.18
	Monopolistic Credit Market	
θ	CES elasticity	2.2
$\frac{\epsilon}{\theta}$	Superelasticity	0.115
	Idiosyncratic Bank Return Risk	
κ	Fraction of portfolio exposed to idiosyncratic risk	0.3
ρ_ξ	Serial correlation of idiosyncratic risk	0.529
σ_ξ	SD of idiosyncratic risk	0.103
	Costly Bank Default	
d_1	Default cost constant	0.0511
d_2	Default cost linear	0.0075

Parameters of the monopolistic credit market block are chosen to hit two targets. First, based on the empirical evidence in Corbae and D'Erasmo (2020a) and Pasqualini (2021), the median markup of commercial banks in the U.S. as of 2020 was roughly 1.8, i.e. 180% over the marginal cost.[8] The average elasticity of $\theta = 2.2$ achieves the CES markup of 1.83. Second, the superelasticity $\frac{\epsilon}{\theta} = 0.115$ generates the variable markup function as seen in Figure 3. In the model, the 95th percentile of markups is roughly 2.6. In the data, it is approximately 3 (Corbae and D'Erasmo, 2020a). The tradeoff here is that having a larger superelasticity increases the maximum of the markup distribution but also imposes a tighter mechanical limit on bank assets. Because the elasticity may never drop below unity, that maximum is $\theta^{\frac{\theta}{\epsilon}}$ and increases in ϵ. Our chosen value ϵ engineers a realistic distribution of markups while at the same time ensuring the mechanical

[8]Using an approach that does not rely on production functions, Jamilov (2020) estimates branch-level loan price and quantity elasticities with respect to instrumented shocks to local credit demand. The author finds that the average nationwide elasticity is 1.2, yielding a large average markup of 6.

constraint is never binding, allowing us to get a reasonably right-skewed banking industry.

Parameters from the idiosyncratic return shocks block are chosen in order to match three facts. First, we motivate κ as the portfolio share that banks allocate to the risky, shadow banking activities. Prior to the Great Financial Crisis the share of shadow banking activities in the broader financial intermediation sector of the U.S. was roughly 1/3 (Gorton and Metrick, 2010). Second, σ_ξ is chosen in order to get the average probability of involuntary exit in line with the data. There are two empirical approaches that we employ for the imputation of that probability. I discuss each of them detail below.

The first approach relies on bank balance sheet and income statement data from the U.S. Call Reports. This database has all the essential data on American commercial banks for the past four decades. We construct an indicator variable Exit_{it} for each bank i and quarter t and run the following logit regression:

$$\Pr\{\text{Exit}_{it} = 1\} = \alpha_0 + \alpha_1 \text{Log (Assets)}_{it} + \alpha_2 \text{Log (Equity)}_{it} + \mu_t \tag{25}$$

Where μ_t is a quarter fixed effect. It captures the idea that the likelihood of exit is heavily aggregate state-dependent. Standard errors are robust to heteroskedasticity and serial correlation. Table 5 describes how we define assets and equity in detail. When both assets and equity are held to their mean value, the predicted probability of exit is 0.974%. We can also compute the probability conditional on different equity values. When assets are held at the average value, the predicted probability of exit ranges from roughly 25% for the first percentile of the equity distribution to almost 0% for banks in the top percentile. Panel (d) of Figure 7 plots the margins plot from our logit regression and depicts all the point and interval estimates of the conditional probability.

The second approach for estimating the probability of bank exit relies on the Laeven and Valencia (2018) database of banking crises from around the Globe. According to the authors, there have been 107 unique banking crises events over the past 48 years across 165 countries.[9] That gives an unconditional probability of 1.35%, which is close to the 0.974% that we compute from the Call Reports and using U.S. data only. As a result, we set σ_ξ such that the average probability of bank default (involuntary exit) in the model is approximately 1% annualized.

Persistence of idiosyncratic bank shocks ρ_ξ is chosen in order to get the skewness of various banking characteristics in the right ballpark. ρ_ξ is a key parameter in the model that directly impacts concentration in the banking sector. A large ρ_ξ (e.g. 0.99) achieves a high degree of concentration and a very right-skewed skewed distribution of leverage (defined as assets over equity), which brings the model closer to the data. On the other hand, Galaasen et al. (2020) find that idiosyncratic borrower shocks that impact bank portfolio outcomes are volatile but not autocorrelated. We

[9] I focus exclusively on incidents of banking crises only, excluding concurring sovereign or exchange rate crises.

therefore set ρ_ξ to 0.529 which is a compromise between empirical evidence on the persistence of idiosyncratic credit shocks and the ability of the model to match banking distributions perfectly.

Finally, the approach to calibrate the real costs of bank default relies on the empirical findings in Laeven and Valencia (2018). Authors estimate that output losses around systemic banking crises historically average around 7.6% of the difference of potential and actual GDP per year. They also find that these losses tend to be larger in advanced economies, which are more financially sophisticated, than in emerging countries. I use this evidence to motivate that default of larger intermediaries in the model is more costly ex-post. I assume that default of a bank in the 90th percentile of the assets distribution in the model corresponds to a banking crisis in a developed economy as defined in Laeven and Valencia (2018) using the World Bank methodology. Similarly, default of a bank in the 10th percentile of the model distribution corresponds to a crisis in an emerging economy. I use a polynomial of order one to map each bank's size to the cost of default:

$$\text{Default Cost(j)} = d_1 + d_2 k(j) \tag{26}$$

where d_1 and d_2 are set to 0.0511 and 0.0075, respectively. These parameters help match the average output loss of 7.6% and the distribution of losses that range from 4.67% to 11.67% in the data.

3.2 Validation

Table 2 lists all key banking moments in the data and in the model. Again, Table 5 describes how we construct and define each variable or ratio. We start with the capital and leverage ratios. Our capital and leverage ratios are slightly greater and lower than in the data, respectively. The reason for this is the following mechanism. The presence of idiosyncratic bank return shocks creates a powerful precautionary lending motive for banks. Banks in the model are effectively risk averse, because the household is, and are thus rushing to outgrow the leverage constraint and the positive default risk region as soon as possible. This leads to a rapid accumulation of net worth. Interestingly, this implies that the riskier the economy is exogenously, the less risky it can become endogenously. This relationship arises in various setups, such as in Fostel and Geanakoplos (2008). Exogenous constraints on the precautionary lending motive, such a lower bound on the deposit rate or additional lending adjustment costs could potentially help solve the issue.

The model does a good job at matching bank markups, both in terms of the average and the 95th percentile. This accomplishment is to a large extend due to the power and flexibility of the Kimball aggregator. The real costs of default, including the mean and the 95th percentile of the distribution, match the data well. The average probability of bank default matches the data almost perfectly. The 95th percentile of default risk ls lower than in the data; this issue is related to

Table 2: **Moments**

	Model	Data		Model	Data
Capital ratio			Markups		
Mean	15.40%	10.80%	Mean	1.81	1.80
95%	28.00%	17.60%	95%	2.58	3.00
Book leverage			Deposit expenses / deposits		
Mean	6.48	11.12	Mean	2.81%	3.75%
95%	8.50	19.20	95%	3.18%	7.30%
Probability of default			Non-interest expenses / assets		
Mean	1.04%	0.95%	Mean	2.15%	0.94%
95%	4.72%	24.29%	95%	2.60%	1.42%
Real cost of bank default			Net interest margin / assets		
Mean	7.78%	7.60%	Mean	1.85%	3.90%
90%	8.81%	11.67%	95%	7.95%	5.40%

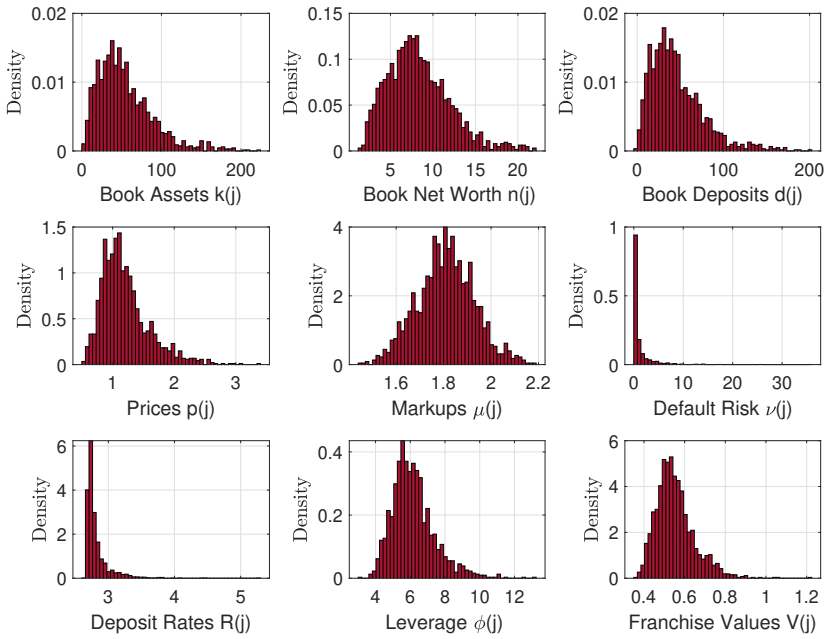
Notes: key moments in the model and in the data. Variables are defined in Table 5. Probabilities are annualized. Markups are absolute. Data source: Call Reports. Data source: U.S. Call Reports. Sample is restricted to U.S. commercial banks. Quarterly data is over 1976:q1-2020:q1.

our inability to generate sufficiently large leverage ratios precisely where it matters - in the left tail of the distribution. In the cross-section of the model, leverage is the best predictor of default risk; the greater it is, the larger the likelihood of involuntary exit. Since capital ratios are too high, endogenous risk is too low, and thus the riskiest intermediaries in the economy are not risky enough.

The ratio of deposit expenses to total deposits is our imputed measure of the ex-post rate of return on deposits ("interest rate"). The mean rate of 2.81% is close to the empirical equivalent of 3.75%. Notice how the value is above the theoretical risk-free rate of 2.60%: the 0.21bps spread is compensation for liquidity and default risk premia. The 95th percentile of the deposit expense ratio is too low, which is not surprising. This is the side effect of the same mechanism that we described above - the riskiest banks are not risky enough. When deposit insurance is turned off, as is the case in the baseline, deposit rates price the distribution of default risk competitively. If default risk is not concentrated enough, the same would happen to the price of that risk.

Non-interest expenses as a fraction of assets are in the right ballpark but are slightly large. This is mainly because we want the problem of the incumbent intermediary to be sufficiently concave all the while maintaining increasing returns to scale. Lowering ζ_2 does not affect any economic channel in the model but worsens its quantitative performance. Finally, the net interest margin to

Figure 6: **Stationary Distributions in the Model**

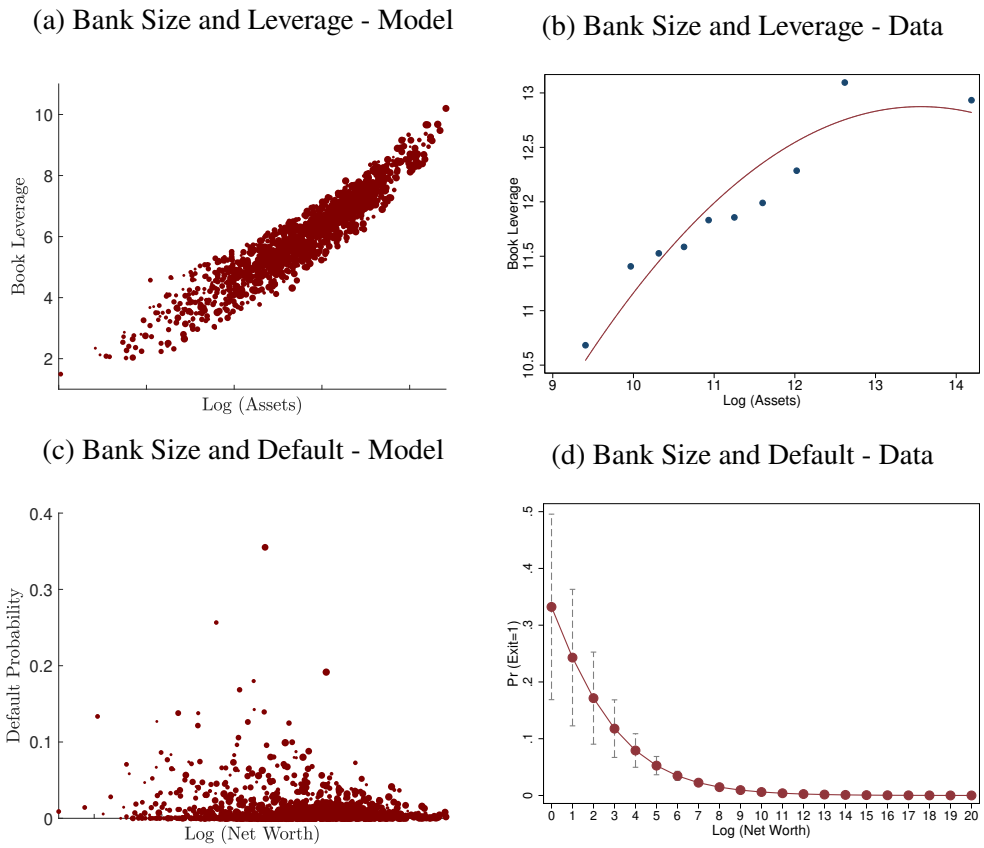


Notes: Model-generated stationary distributions.

assets ratio is in line with the data.

We now look at all stationary cross-sectional distributions that the model generates. Figure 6 plots univariate histograms for bank assets k(j), net worth n(j), deposits d(j), market leverage ϕ(j) relative prices p(j), markups μ(j), default risk ν(j), deposit rates $\bar{R}$(j), and franchise values V(j). In line with the data, the credit market is concentrated, i.e. there is a small fraction of large and profitable intermediaries with a significant market share of assets, deposits, and net worth. The distribution of markups has the same right-skewed shape as in, for example, Pasqualini (2021): the right tail is driven by the largest banks in the economy who charge the highest markups. Distributions of default risk and deposit rates, which feed into relative prices through the marginal cost channel, are of a similar right-skewed and dispersed shape. Here, in contrast, the right tail is driven by a small share of very low-net-worth intermediaries with risky balance sheets and high marginal costs.

Figure 7: **Cross-Sectional Patterns in Model and Data**



Notes: Cross-sectional relationships between measures of bank size, default risk, and book leverage, in the model and data. In Panel (b) the x-axis is binned into 10 deciles of the total assets distribution. For each bin, book leverage is computed as the ratio of bin-specific total assets over bin-specific total equity. In Panel (d), marginal point and interval estimates are from the logit regression of the indicator for exit on log (assets) and log (equity) with a time fixed effect and standard errors that are robust to serial correlation and heteroskedasticity. The margins plot shows conditional probabilities of exit when log (assets) are held at mean value. Panels (a) and (c) show regular scatter plots based on a stochastic simulation with N=1 intermediaries and T=2,000 quarters. Data source: U.S. Call Reports. Sample is restricted to U.S. commercial banks. Quarterly data is over 1976:q1-2020:q1.

We now focus on two essential cross-sectional patterns from the data that the model matches. They are important because together they constitute the policy trade-offs which we define and discuss in the next sections. First, in line with the data, the model generates a positive cross-sectional correlation between bank book leverage and size. Figure 7 visualizes the result in Panel (a) for data and (b) for model. We proxy bank size with total assets, similarly to Adrian and Shin (2010) and Coimbra and Rey (2019), and continue with the same sample of banks from the Call Reports. In order to minimize the influence of outliers and noise in the data, we plot binned scatter plots in Panel (b). Specifically, we construct ten deciles (bins) of the pooled distribution of total assets across 1976:q1-2020q1. For each bin, we compute the ratio of bin-specific total assets over bin-specific total equity. Both in book values. For the model in Panel (a), we plot a regular scatter

plot from a stochastic simulation that includes N=1 intermediaries and T=2,000 quarters. The positive association between assets and leverage can be clearly seen on both plots.

Second, the model correctly predicts that the distribution of default risk is concentrated in the left tail of the bank net worth distribution. We already discussed how we define and estimate the empirical likelihood of exit in the previous paragraph. In the model, the object of interest is the usual probability of insolvency $\nu(j)$. Figure 7, Panels (c) and (d) shows how the relationship between size and exit risk looks in the model and in the data - it is negative in both cases. In both Panels, it can be seen that larger-than-median intermediaries are essentially risk-free. In the lower deciles, however, exit risk escalates rapidly and goes beyond 10% both in the data and in the model.[10]

4 Bank Regulation

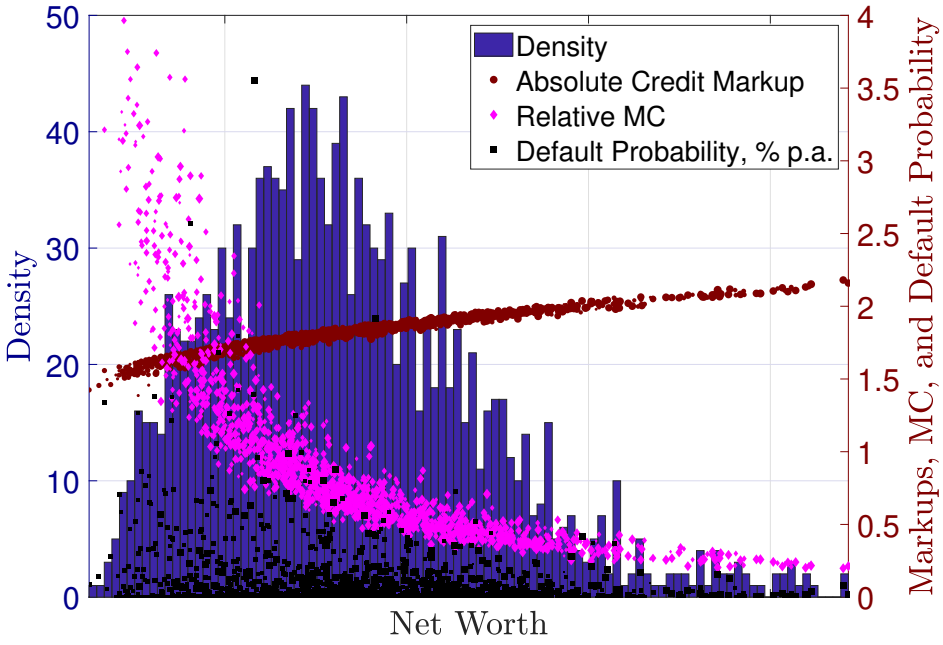
In this section I first define the tradeoff between bank competition, stability, and efficiency. I then discuss how introducing heterogeneous capital requirements or deposit insurance affects the macroeconomy through the prism of this tradeoff. I conclude by solving for constrained-efficient allocations of a social planner.

4.1 The Competition, Efficiency, and Stability Trilemma

The tradeoff between financial competition, efficiency, and stability arises due to the interaction of three channels: economies of scale, size-markup complementarity, and costly default. Figure 8 visualizes the mechanism. The picture plots the stationary distribution of bank net worth in the background. Overlayed are scatterplots for absolute markups $\mu(j)$, relative marginal costs, and absolute probability of default $\nu(j)$ in percent p.a. The economies of scale channel is represented by the negative relationship between marginal costs and net worth - larger banks are more cost-efficient and have a greater marginal propensity to lend MPL(j). The size-markup complementarity channel is seen from the positive relationship between markups and net worth. Finally, the default risk channel is seen from the negative relationship between the default probability and net worth. The trilemma exists because banks that are efficient and stable are the same ones that charge higher markups. Efficiency gains from having more banks with low $\nu(j)$ and high MPL(j) is counteracted by them contributing to a higher average markup and, as a result, greater welfare losses from bank

[10]For completeness, it is important to mention that the model-based bank size distribution is not nearly concentrated enough. Idiosyncratic risk is not enough to generate "superstar banks" whose presence we observe in the data. This issue has been a challenge for the literature on household wealth inequality for years. One remedy to this problem could be the introduction of ex-ante heterogeneity in balance sheet growth types as in Gabaix et al. (2016). I leave this for task for future research.

Figure 8: **The Banking Industry Trilemma**



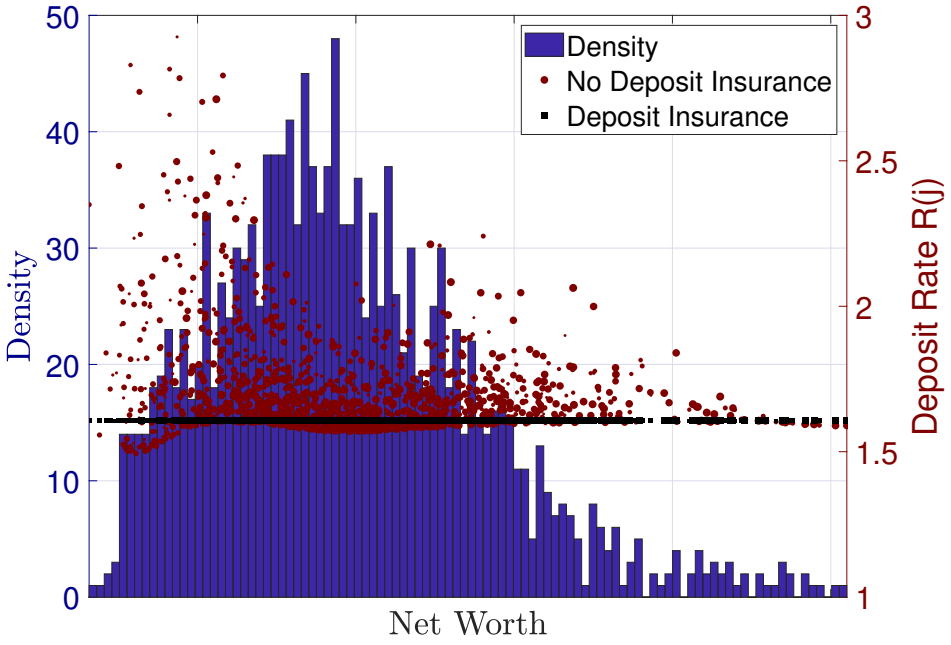
Notes: distribution of bank net worth and scatter plots of markups, relative marginal costs, and probabilities of default.

market power.

At the heart of the trilemma are two intertwined *bilateral* tradeoffs. First, the canonical competition-stability tradeoff. Monopolistic competition allows high-net-worth banks to reach greater equilibrium franchise values through higher markups. This, in turn, reduces appetite for private risk-taking, lowering the probability of insolvency in equilibrium. Second, the competition-efficiency tradeoff. High-net-worth intermediaries charge higher markups but they are also more efficient from the cost- and productivity standpoints, as discussed previously. Each of the two bilateral tradeoffs can be viewed to rely on the variable markups channel. With constant markups, we may still entertain an efficiency-stability tradeoff if the physical cost of default is sufficiently convex in bank size. This way, even though bigger banks default less often, realized social costs of their rare defaults would be able to counteract, in expectation, the efficiency gains from economies of scale.

It is crucial to emphasize that the trilemma does not imply that it is *impossible* for the regulator to improve *net* welfare. Net quantitative effects always depend on calibration of the credit demand superelasticity, the cost of default function, and of the idiosyncratic risk process. Suppose we consider an economy where the largest intermediaries are state-run. Physical costs of default of banks in the right tail would therefore potentially always outweigh any efficiency losses from high

Figure 9: **Deposit Insurance Scheme**



Notes: distribution of bank net worth and scatter plots for deposit rates in the economy with deposit insurance. Black squares represent equilibrium rates when guarantees are turned on and red circles the counterfactual rates if guarantees were turned off.

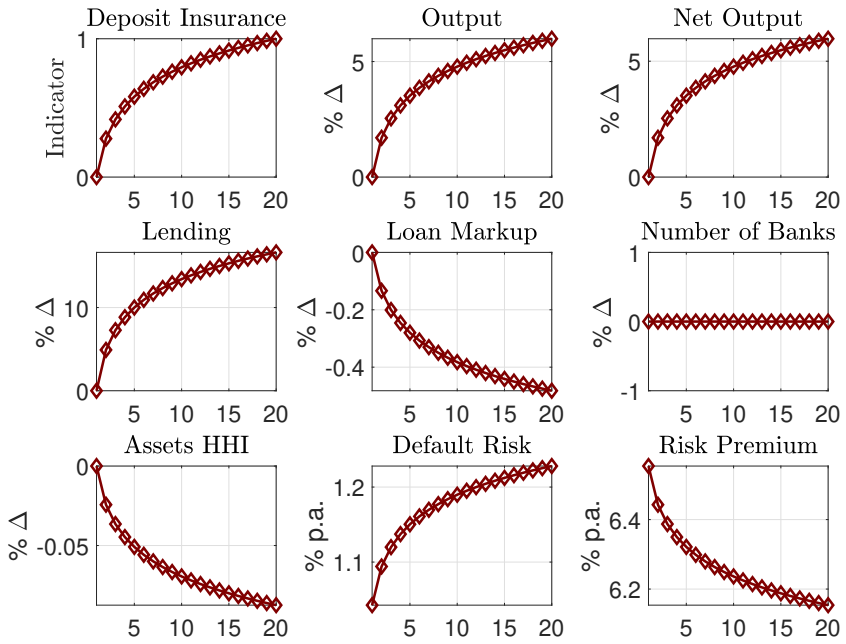
markups in "normal times". We are merely claiming that if a regulator attempts to improve any side of the trilemma, one or all of the remaining two dimensions would necessarily deteriorate as a matter of unintended consequences. Whether the policy shock or the unintended side-effects dominate is a quantitative question.

4.2 Deposit Insurance

We now explore how in our calibrated model various policy schemes affect the macroeconomy through the prism of the aforementioned trilemma. We begin with deposit insurance. When deposit guarantees are switched on, the distribution of equilibrium deposit rates $\bar{R}(j)$ is flat. To achieve this as an endogenous result, banks continue to take as given their individual default probability $\nu(j)$, but we break the mapping between $\nu(j)$ and $\bar{R}(j)$ and resolve the model. In other words, there is no equilibrium pass-through from balance sheet riskiness to the prices of debt. The government promises to honour any deposit shortfalls due to endogenous bank exit. We assume that the government funds the scheme via lump-sum taxes on the household.

Figure 24 shows the outcome of this policy. In the background is the new stationary distribution

Figure 10: **Macroeconomic Effects of Deposit Insurance**



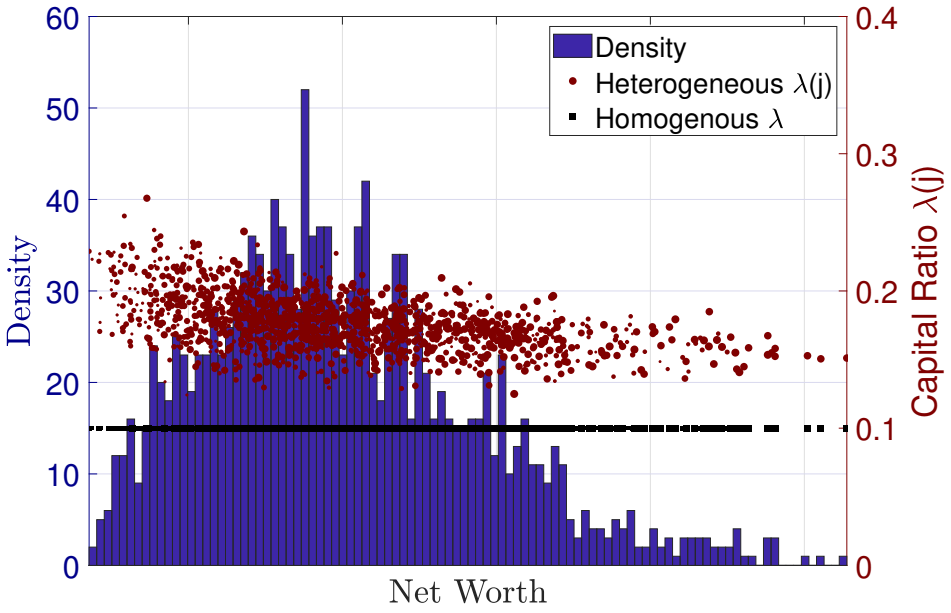
Notes: macroeconomic effects of switching on deposit guarantees. Net output is defined as output Y_t net of real costs of bank default.

of bank net worth which is consistent with the equilibrium with deposit insurance. The flat-lined black scatter plot shows the invariance of the equilibrium $\bar{R}(j)$ with respect to bank size. For contrast, the red scatter plot represents the counterfactual rates if guarantees were turned off; notice the usual inverse relationship with n(j) in that case. It can be seen from the Figure that the biggest beneficiaries from the introduction of deposit guarantees are low-net-worth banks with high marginal costs.

Figure 10 demonstrates the macroeconomic effects of deposit insurance. At t=0, we start from the baseline stationary industry market equilibrium. At t=20, the economy has permanently converged to its version with full deposit insurance. Lending, output, and net output (net of realized costs of bank default) all increase since funds are now cheaper to obtain. In line with most theoretical and empirical evidence on the interactions between risk-taking and deposit insurance, we see a positive increase on average default risk. Interestingly, we find that the average markup falls. This is consistent with the fact that the deposit insurance tool favors small banks by more. The market share of of low-net-worth banks increases, concentration (Assets HHI) falls, and the average markup falls.

The mechanism of the trilemma holds - introducing deposit insurance has increased lending and output but also raised systemic riskiness. Quantitatively, net output has still grown because

Figure 11: **Heterogeneous Capital Requirements**



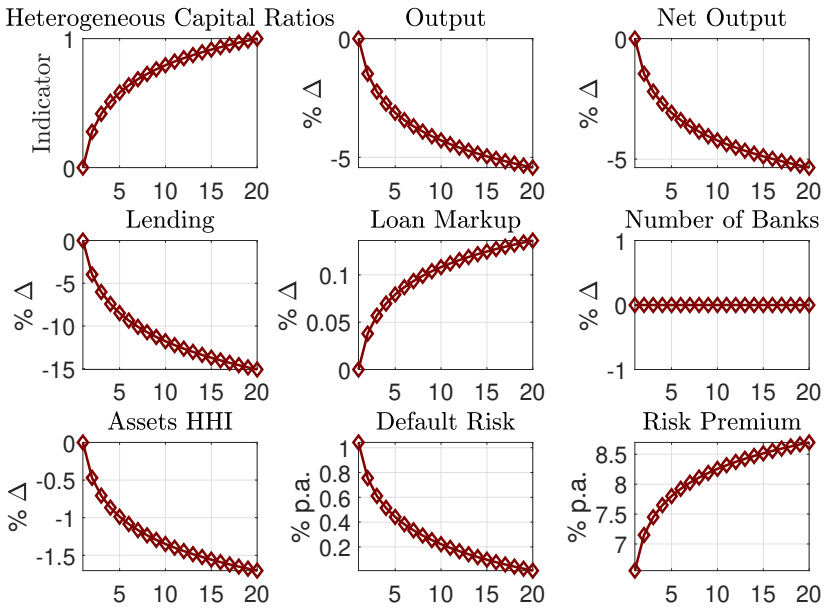
Notes: distribution of bank net worth and scatter plots for λ under homogenous and heterogeneous capital requirement regimes.

default costs turn out to be negligble. Recall that the number of banks is time invariant because for now we still operate with an exogenous entry margin. Finally, aggregate risk premium, defined as $R^k - \bar{R}$ (annualized return on aggregate capital minus the average deposit rate) falls since aggregate quantities rise and prices fall - both effects lower R^k.

4.3 Heterogeneous Leverage Regulation

The second market intervention that we consider is heterogeneous capital requirements. One possibility to impact private risk-taking and aggregate fragility is to impose regulatory limits on ϕ(j). In practice, this corresponds to micro-prudential regulation which is a common practice by governments around the world. Recall that in the market economy, leverage falls with bank size while λ is homogenous across all banks. We now consider a scenario where λ(j) is ex-ante heterogeneous and falls linearly with bank net worth. Banks in the top decile of the distribution face the same $\lambda_{\iota=10} = 0.1$ as before. However, banks in the lowest decile face $\lambda_{\iota=1} = 0.3$. All banks in the deciles that are in between face a λ(j) that is interpolated based on the exogenous grid of net worth and their position in the distribution. What this policy is designed to achieve is to restrict leverage of precisely those intermediaries who are the most likely to have high leverage to begin with.

Figure 12: **Macroeconomic Effects of Heterogeneous Capital Requirements**

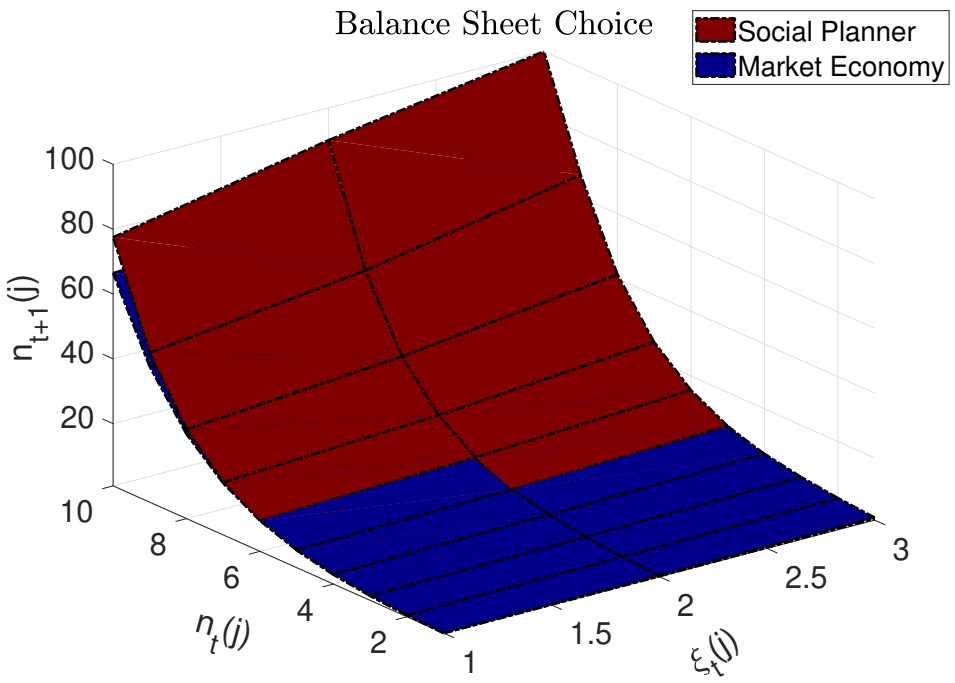


Notes: macroeconomic effects of switching on heterogeneous capital requirements.

Figure 11 shows how the policy works in the model. Overlayed on the new equilibrium distribution of net worth are the homogenous λ from the baseline economy and λ(j) from the economy with capital requirements. The negative slope of the λ(j) scatter plot implies that the policy has achieved its desired objective - limitations on leverage are proportional to actual leverage, here summarized by n(j) as the sufficient summary statistic. Recall that ϕ(j) falls with n(j), as demonstrated and discussed earlier in Figure 5.

Figure 12 portrays the macroeconomic effects of this policy. Similarly to before, t=0 and t=20 correspond to the baseline regime and the case with heterogeneous λ(j), respectively. We see that all aggregate quantities have fallen, including lending, output, and net output. This is driven by the fact that the leverage constraint is now tighter precisely for the agents for which it is more likely to bind - the low-net-worth banks. As a result, aggregate demand for deposits and bank leverage are down. Notice how default risk has fallen by a considerable amount - the average probability of default is almost 0%. The policy is highly successful in terms of reducing systemic financial fragility. The tradeoff here is of course the reduction in efficiency, intermediation, and production. Risk premia are up because aggregate capital is down and prices are (slightly) up. Concentration is slightly down because the new λ(j) regime is detrimental for precautionary lending growth for practically all banks in the economy, with the exception of the top decile. As a result, the distribution is less dispersed and less right-skewed as fewer banks are capable of accumulating a lot of net worth.

Figure 13: **Market Equilibrium and Constrained Efficient Allocations**



Notes: Market-based and social planner's allocations from the stationary equilibrium.

There are fewer "superstar" banks in equilibrium.

4.4 Constrained Efficiency and Optimal Policy

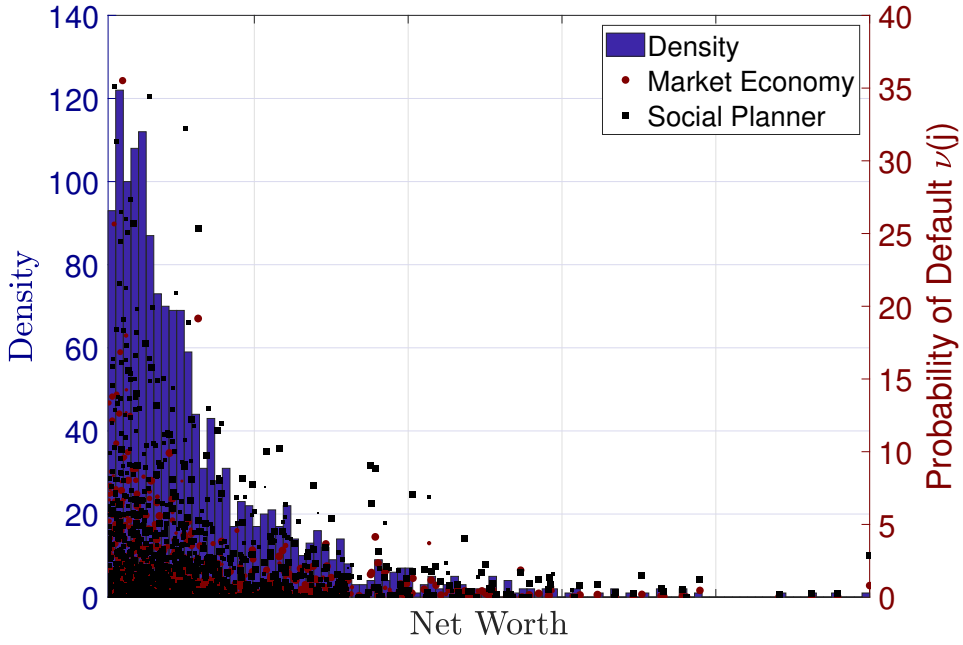
We now consider constrained-efficient allocations of a hypothetical social planner as a stepping stone for optimal policy analysis. The planning problem is identical to that of the baseline economy with one crucial exception. In this section, the planner picks the quadruple $\{k, d, p, \mu\}$ in order to maximize the franchise value V while understanding that R^T is *endogenous* through the impact of the quadruple on R^k and P. Consider the law of motion of net worth that the social planner faces:

$$n_{t+1}(j) = R^T\Big(n(j), \xi(j), \{k_t(j), d_t(j), p_t(j)\}\Big)p_t(j)k_t(j) - \bar{R}_t(j)d_t(j) - \frac{1}{\zeta_1}k_t(j)^{\zeta_2} \tag{27}$$

Compare this formula to Equation 9 from the market equilibrium. The difference is that R^T is no longer taken as given. Numerically, the banking problem is solved under the assumption that R^k and P are both polynomials in $\{n(j), \xi(j), \{k_t(j), d_t(j), p_t(j)\}\}$. We use projection methods to solve for the coefficients that are consistent with equilibrium. See Appendix E for more details on the numerical algorithm.

Figure 13 presents the two-dimensional optimal choice of next-period bank net worth $n'(j)$.

Figure 14: **Systemic Risk Implications of Constrained Efficiency**

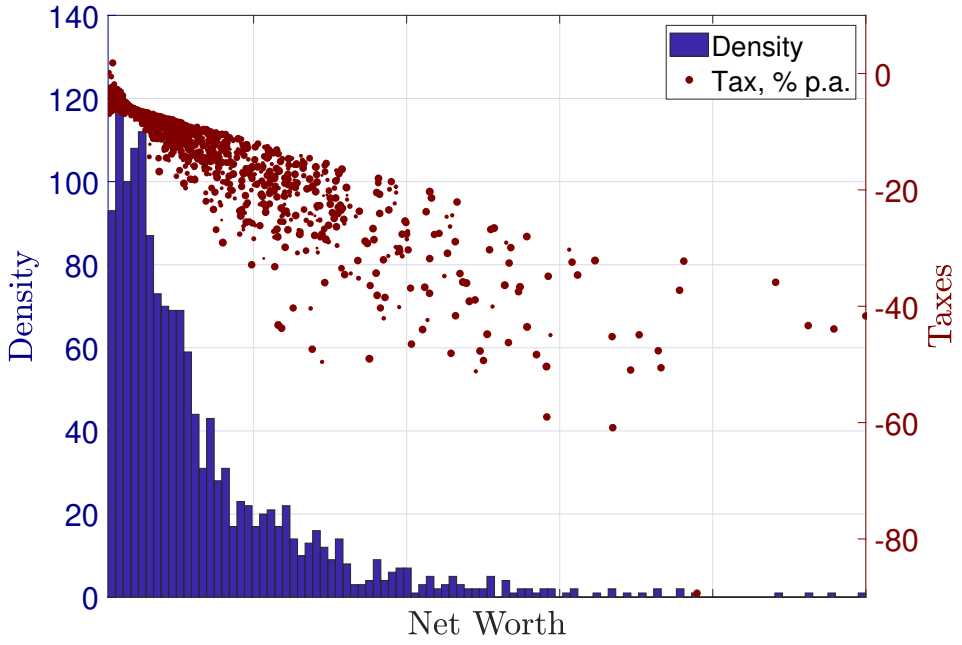


Notes: distribution of bank net worth and scatter plots for $\nu(j)$ under alternative market regimes.

We contrast decisions of the social planner with the market outcome. Comparing the two cases reveals that misallocation is present in the decentralized equilibrium along both the net worth and idiosyncratic risk dimensions. Specifically, the market outcome yields too *little* lending because of an aggregate credit supply externality. Monopolistic credit competition leads to underutilization of risky capital as a resource in production. Unlike the social planner, individual banks do not internalize the impact of their private choices on aggregate returns. In addition, misallocation is more sever for higher levels of net worth. This is consistent with the idea that markups are variable and increase with net worth. Recall that demand for firm claims is more satiated in the right tail of the bank size distribution.

Figure 14 shows how equilibrium financial stability responds to social planner's actions. We continue to define financial stability as the probability of bank default due to insolvency $\nu(j)$. We plot the new stationary distribution of net worth and $\nu(j)$ scatter plots that correspond to the constrained efficient (blue square) and market (red circle) allocations. Here we observe that the social planner's solution induces a considerable increase in system-wide default risk. Low-n(j) intermediaries become particularly more risky. This result is a case in point of the financial stability-competition trade-off (Hellman et al., 2000). Specifically, the social planner targets the credit supply externality by reallocating resources towards agents with the highest marginal propensity to lend - the bigger

Figure 15: **Optimal Policy**



Notes: distribution of bank net worth and scatter plots for the optimal $\tau(j)$, in percent p.a.

banks. However, smaller intermediaries are fundamentally more prone to insolvency risk to begin with. Small, risky banks become relatively riskier. Average probability of default, as a result, goes up and the economy is more fragile.

We decentralize constrained efficient allocations with taxes on banks gross returns. Importantly, these policies are size- and income-dependent because misallocation and markups correlate with the joint distribution of bank net worth and idiosyncratic risk. Theoretically, gross returns taxes are easier to operationalize because they target specifically the wedge in the bank-specific total portfolio return process and the law of motion of net worth. Specifically, we conjecture a size and idiosyncratic return specific tax rule $\tau\Big(n(j), \xi(j)\Big)$ and impose it on the market equilibrium. Computationally, we assume that taxes are polynomials in $n(j)$ and $\xi(j)$ and solve for coefficients that are consistent with a minimal distance between the equilibrium and the social planner allocations. See Appendix E for details. Note that negative taxes (subsidies) are allowed, which is important when working with underutilization of resources due to monopolistic competition. The law of motion of bank net worth with tax policies is now:

$$n_{t+1}(j) = R_t^T(j)\left[1 - \tau\Big(n(j), \xi(j)\Big)\right] p_t(j)k_t(j) - \bar{R}_t(j)d_t(j) - \frac{1}{\zeta_1}k_t(j)^{\zeta_2} \tag{28}$$

Table 3: **Summary of Allocations**

	Market Economy	Deposit Insurance	Capital Require-ments	Optimal Policy
Output	4.204	4.463	3.981	4.822
Net Output	4.200	4.459	3.981	4.813
Book Leverage	6.170	6.255	4.492	6.270
Systemic Risk	1.04%	1.23%	0.01%	2.24%
Aggregate Markup	1.821	1.813	1.824	1.830

Notes: Macroeconomic and financial aggregates across different regulatory and market regimes.

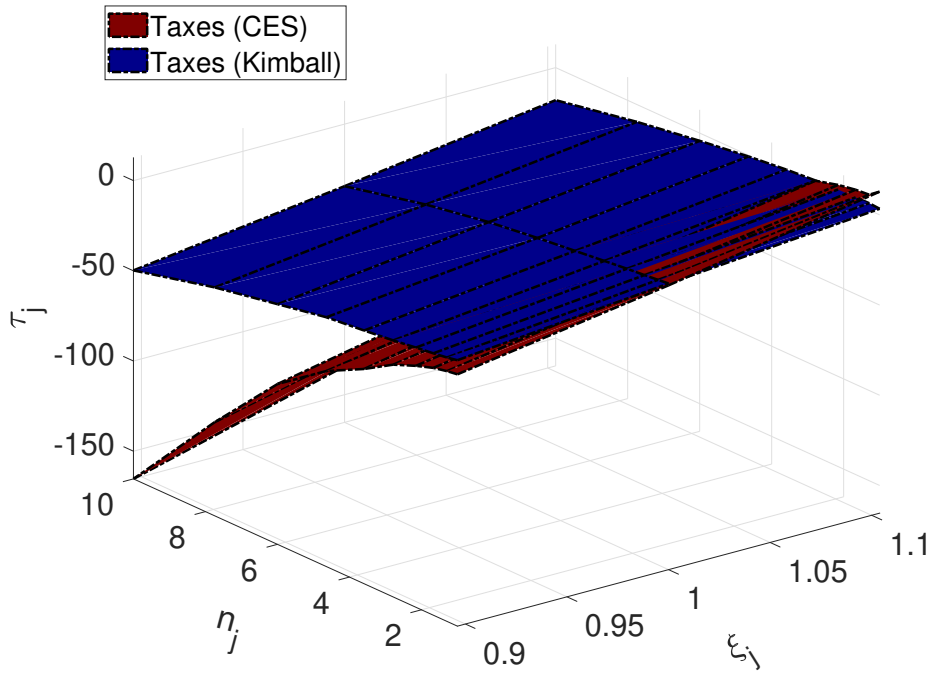
Effectively, on each point in the grid, we search for tax values that equalize socially optimal and market allocations.

Figure 15 plots the stationary distribution of net worth from the social planner's problem with the scatter plot for optimal taxes $\tau(j)$. Notice how the vast majority of intermediaries in the state space receive a *subsidy*. The subsidy is the highest (in absolute terms) for *big* banks. On the other hand, low-net worth intermediaries may even sometimes face a tax. The intuition for this result is related to the analysis of constrained efficiency: marginal propensity to lend (MPL) increases with bank net worth. The social planner finds it most efficient to correct the under-lending externality by stimulating/subsidizing lending of those with the highest marginal propensity to respond to taxes. In general equilibrium, this increases aggregate output and household consumption. In the stationary distribution, the annualized tax ranges from -60% to 2% with the average tax of about -12.32% per year.

To summarize our findings across different regulatory regimes and market structures, we report all key aggregates in Table 3. We focus on output, net output, bank leverage, systemic risk, and the aggregate markup. Output is aggregate production from the stationary steady state. Net output is gross output adjusted for the real costs of realized bank default. Book leverage is the unweighted average of $\frac{k(j)}{n(j)}$. Systemic risk is defined as the average probability of bank default $\nu(j)$. The aggregate markup is the unweighted average of $\mu(j)$.

We start with the first column - the market economy. Introduction of deposit insurance raises aggregate output at the cost of greater leverage and systemic risk. Heterogeneous capital requirements, on the other hand, virtually eliminate financial instability but reduce aggregate efficiency. Optimal policy, i.e. heterogeneous taxation of bank portfolio returns, achieves the highest possible level of output which is not surprising considering this is the second-best outcome. But it also leads to the highest probability of bank default among all considered cases. Notice how in all scenarios, net output is quantitatively indistinguishable from gross output. This is because the cost

Figure 16: **Optimal Taxes under CES and Kimball Aggregators**



Notes: optimal taxes τ(j) under constant (CES) and variable (Kimball) markups.

of default is not sufficiently large enough on average and that the likelihood of default is the highest for small-n(j) banks, whose real cost of default is relatively smaller.

The Role of the Aggregator An interesting auxiliary exercise is to compare normative implications under the two alternative regimes for bank markups: constant and variable. Figure 16 plots the two-dimensional policy function for optimal bank taxes under the Kimball and CES aggregators. We see that the CES economy implies much greater subsidies on average and to high-n(j) banks in particular. The reasoning behind this result is related to Edmond et al. (2018) who make a similar point in a model with variable markups and non-financial firms. The CES model is essentially "misspecified" if the true model features variable markups. The CES aggregator mistakes lack of lending from bigger banks for an opportunity to enhance welfare. In practice, however, if the true elasticity is declining with n(j), then the additional resources would run into the problem of diminishing marginal values for new claims.

5 Applications

In this section I explore several applications of the framework and of the banking trilemma. We explore the predictions of the framework for the ongoing global rise in banking concentration and the emergence of fintech-intermediated credit. We proceed by examining the too-big-to-fail hazard and implementation of various targeted stabilization policies. Finally, we conclude with implications for intermediary asset pricing models and empirics. Appendix C.1 considers additional targeted policies such as bank-level lending facilities and debt guarantee schemes. In Appendix D we study crisis events by simulating MIT shocks to aggregate productivity.

5.1 The Rise of Banking Concentration

The banking industry around the world is becoming more and more concentrated (Corbae and D'Erasmo, 2020b; Constancio, 2016). We do not take a stance on the *cause* behind the rise of concentration. Instead, we quantify the impact of various distributions of banks on the macroeconomy by fitting several counterfactual cross-sectional distributions of bank assets into the stationary general equilibrium and re-evaluating all endogenous variables that would be consistent with them. Counterfactual distributions are generated exogenously by drawing sequences of bank assets $k_1 \dots k_N$ from well-known continuous probability densities such as Uniform or Pareto. We fit each generated sequence into the model, re-compute all policy functions, but do not run the step which calculates new distributions. In other words, we solve for partial-equilibrium policy functions that are consistent with the exogenously constructed distributions.

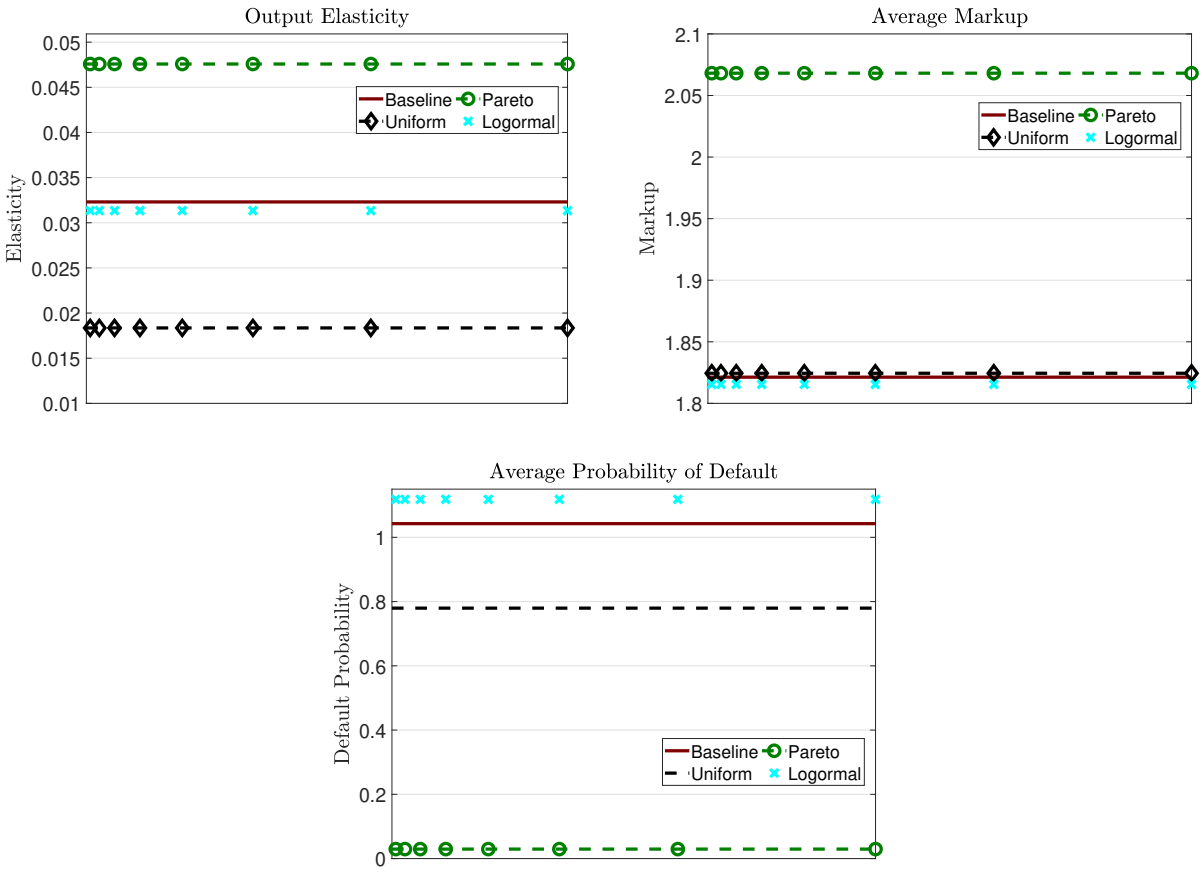
We consider 3 broad families of densities: Uniform, Lognormal, and Pareto. For the uniform density, we generate N=20,000 random numbers from the interval $[0.5K_{ss}, 1.5K_{ss}]$, where K_{ss} stands for the level of aggregate capital in the market equilibrium. For the lognormal density, we draw from $P(\mu_k, \sigma_k^2)$, where μ_k and σ_k are, respectively, the mean and standard deviation of the k(j) distribution from the stationary equilibrium. For the Pareto density, we follow (Gabaix, 2009) and consider a common case of the power parameter $\alpha = 2$.[11]

We focus on three aggregate variables of interest - the output elasticity of uniform bank net worth shocks, the average markup, and the average probability of default. These three objects summarize the three dimensions of the banking industry trilemma. The output elasticity can be defined with the help of the previously analyzed marginal propensities to lend (MPL):

$$\frac{\partial Y}{\partial N} = \underbrace{\frac{\partial Y}{\partial K}}_{\text{MPK}} \times \int_{\mathbf{B}} \underbrace{\frac{\partial k(j)}{\partial n(j)} \mu(dn, d\xi)}_{\text{MPL(j)}} \tag{29}$$

[11]The scale parameter is chosen to be a factor of k_{min}, i.e. the minimum level of assets from the market economy.

Figure 17: **Macroeconomic Effects of Alternative Banking Distributions**



Notes: Output elasticities, probabilties of default, and aggregate markups across alternative cross-sectional distributions.

where the first term on the right-hand side is the marginal product of capital and the second term is the aggregate MPL. We treat a high output elasticity as a symptom of high efficiency in the lending market.

Figure 17 presents the results of this exercise. We observe that the output elasticity with respect to uniform net worth shocks is highest for the Pareto economy, followed by the baseline, lognormal, and uniform economies. Intuitively, the degree of right-skewness can be viewed as a sufficient statistic for the elasticity, and thus efficiency. Similarly to what we concluded based on Figure 8, the bigger the share of high-n(j) banks the higher aggregate efficiency gets. Because of the economies of scale channel, larger banks have both a greater MPL and a lower marginal propensity to price (MPP). The fact that the Pareto economy, which is more concentrated than our baseline model, has a higher elasticity is proof of the mechanism. The uniform density has the lowest elasticity which numerically corresponds very closely to the elasticity of the representative-bank special case.

From Figure 17 we also see that the average markup is the highest for the Pareto economy, followed by the three other alternatives which are hard to distinguish from each other. The Pareto economy is by far the most concentrated of the four, and its largest banks choose abnormally high credit markups. As a result, the aggregate markup gets very inflated. Finally, the average probability of default is the lowest in the Pareto economy, followed by uniform, baseline, and lognormal economies. The degree of concentration can be viewed as a good predictor of systemic stability: the Pareto economy is the most concentrated and is thus the least risky.

Overall, this exercise is a simple but useful demonstration of quantitative implications of the banking trilemma. The most concentrated economy, whose distribution is drawn from a Pareto density, is the most efficient, least competitive, and most stable. The prediction of our framework for the future of banking is thus the following. If banking concentration continues to go up, which seems to be a realistic assumption to make given the cost-cutting and competitive trends, then the macroeconomy will benefit from higher efficiency stemming from the right tail, will attain a greater buffer against financial crises, but will suffer from welfare losses due to rising financial markups. This prediction seems to be in line with the recent time-series experience of the U.S. banking sector: the industry has become more concentrated all the while markups have risen (Corbae and D'Erasmo, 2020a).

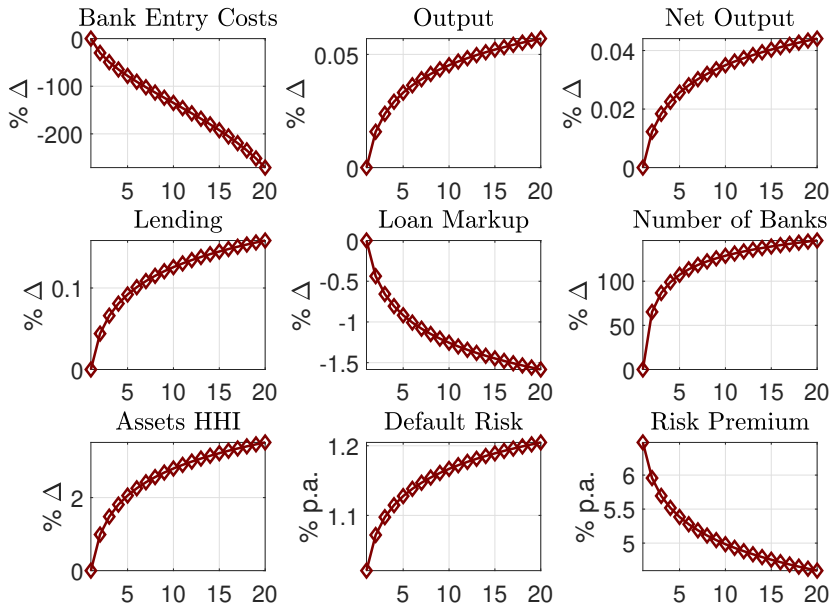
5.2 Emergence of Fintech Credit

The global share of fintech in financial intermediary activities is growing rapidly, both in developed and developing financial markets (Claessens et al., 2018). In order to formalize the rise of fintech/bigtech firms, I extend the baseline model with endogenous bank entry in the spirit of Melitz (2003). There is now infinite mass of aspiring financiers who specialize in banking services. Before entry, every financier pays a fixed entry cost e in units of capital. The rise of fintech will be simulated as a permanent decline in e. This is a reduced-form stand-in for various possible technological and preference-based explanations for this trend. Having paid the sunk cost, the financier receives an idiosyncratic return profitability draw $\xi_0 \in \Xi$ from the ergodic distribution $G_0(\xi)$ that is implied by the ξ process. The financier is also bestowed with an initial level of net worth n_0 which is a constant fraction of the aggregate stock of net worth N. Afterwards, the financier decides whether to operate or to immediately exit. Conditional on its state $\{n_0, \xi_0\}$, the financier operates if and only if its expected discounted franchise value exceeds e. The value function of the entering financier is therefore:

$$V^e(n_0, \xi_0) \equiv \max \left[V(n_0, \xi_0) - e, 0 \right] \tag{30}$$

Free entry drives the future expected excess value of the entering intermediaries, net of startup

Figure 18: **Fintech Credit Growth**



Notes: Simulation of the fall in bank entry costs in the economy with endogenous entry.

costs, to 0. A financier's incentive to enter is driven by the desire to earn economic profit. Entry keeps occurring until expected bank profits are equalized with the cost of financial variety origination. In equilibrium, either V^e is equal to 0, the number of entrants is 0, or both.

The mass of financiers that decide to enter is M. The mass of active intermediaries, which now includes both incumbents and new entrants, is H. The stationary distribution of banks now keeps track of M as well as the incumbents:

$$\mu'(n', \xi') = \underbrace{\left(\sigma - E'\right) \sum_{\xi} G(\xi', \xi) \int \mathbb{1}_{\{(n,\xi)|K(n,\xi)\in \mathbf{B}\}} \mu(dn, d\xi)}_{\text{Surviving Incumbents}} + \underbrace{M' \int \mathbb{1}_{\{(n_0,\xi)|K(n,\xi)\in \mathbf{B}\}} G_0(\xi)}_{\text{New Entrants}} \tag{31}$$

The law of motion of the distribution is now:

$$\eta_{t+1}(n_{t+1}, \xi_{t+1}) = \Phi(\eta_t, M_{t+1}) \tag{32}$$

Credit market clearing now requires aggregate supply to equal the demand from the incumbents and the entrants:

$$\underbrace{K}_{\text{Aggregate Supply}} = \underbrace{\int_{\mathbf{B}} \Big(k(n,\xi)\Big)\mu(dn,d\xi)}_{\text{Incumbent Demand}} + \underbrace{M\int_{\mathbf{B}} \Big(k(n_0,\xi_0)\Big)dG(\xi_0)}_{\text{Entrants Demand}} + \underbrace{Me}_{\text{Entry Cost}} \tag{33}$$

We set $e = 1.65$ for the baseline case. The fintech economy has $e = 0.11$. The number is calibrated such that the number of active banks in the economy roughly doubles.

Figure 18 shows the result of this exercise in the usual format. The model predicts that fintech credit will be responsible for more lending, output, and the number of active intermediaries - this is the direct extensive margin effect. Because the average intermediary is smaller, this lowers the average markup. We also observe a considerable elevation in financial fragility. Low entry costs essentially allow "too many" low-type lenders to enter every period by lowering the minimum profitability threshold below which financiers do not wish to stay. A growing mass of low-size, high-risk young intermediaries contributes to rising systemic fragility since default risk falls with net worth. This prediction is in line with the belief among regulators and policy-markers that fintech credit is a major source of financial stability for the 21st century.

5.3 Too Big to Fail

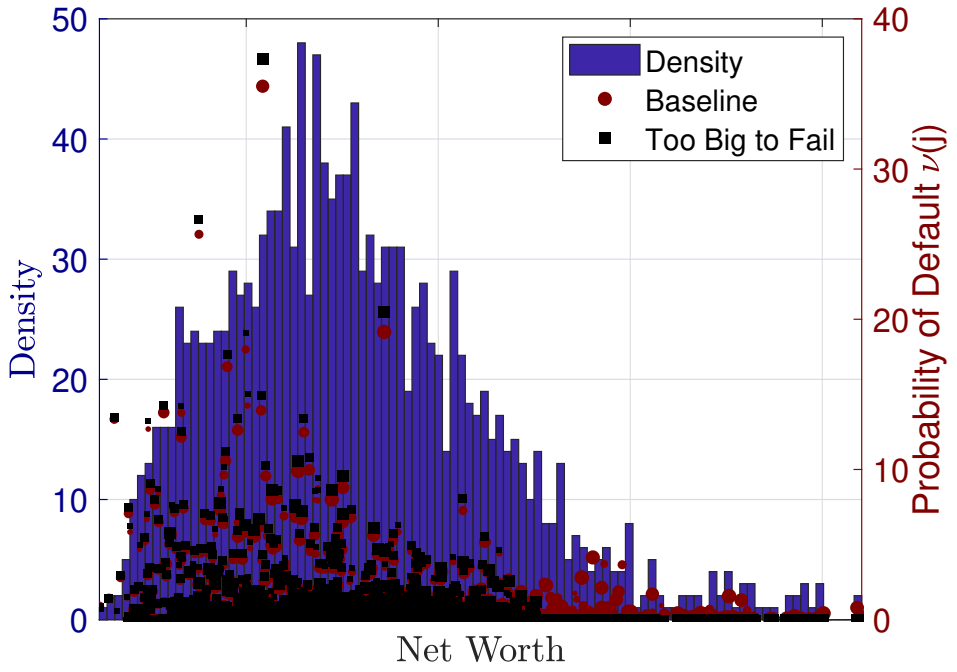
Absence of effective bank failure-resolution rules and laws pre-Lehman meant that systemically important banks, particularly those with U.S. headquarters, benefited from implicit "too-big-to-fail" (TBTF) subsidies. Probability of an ex-post government bailout of large financial institutions was close to one, which was priced by the market into lower debt financial costs after the adjustment for insolvency and illiquidity risks. Conditional on this safety net being part of the environment, market participants lose their incentive to monitor the intermediaries and banks lose the incentive to act prudently, which further exacerbates the problem (Stern and Feldman, 2004).[12]

In the context of our model, we operationalize the TBTF hazard problem the following way. The probability of default $\nu(j)$ of any bank in the top decile of the distribution of net worth is zero, regardless of balance sheet properties. We pick the top decile simply for quantitative tractability. The bottom nine deciles instead face a $\nu(j)$ that is consistent with their size-risk profile as usual. The policy function for $\nu(j)$ is therefore kinked, and all banks understand this. The pass-through from $\nu(j)$ to $\bar{R}(j)$ functions normally - all banks face the cost of debt that is consistent with their probability of default. This implies that some banks will face an exogenously imposed "cost of funds subsidy" which switches on only if the bank reaches a certain size threshold.

Figure 19 shows how the mechanism works. Observe how the scatter plot for $\nu(j)$ in the TBTF case is clearly kinked: banks in the right tail of the distribution face no default risk exogenously.

[12] In recent years, multiple studies have found that the TBTF problem has declined after the passage of the Dodd-Frank Act. (Haldane, 2010; Atkeson et al., 2018)

Figure 19: **TBTF**

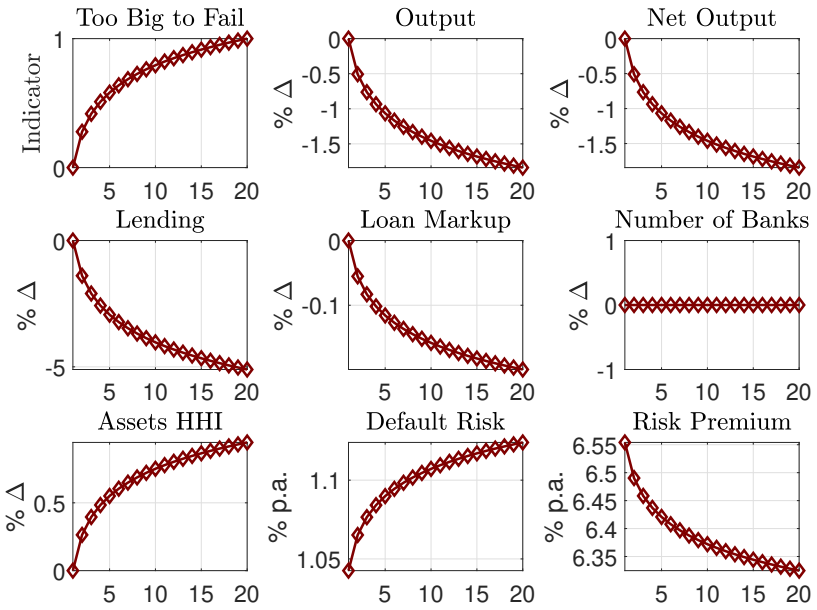


Notes: distribution of bank net worth and scatter plots for $\nu(j)$ for the baseline economy with and without the TBTF subsidy.

In contrast, in the baseline economy some of the same banks in the right tail face a positive $\nu(j)$. In addition, it is interesting that most banks in the bottom nine deciles now face a *higher* $\nu(j)$. The TBTF subsidy, even if it affects only the largest intermediaries, makes leverage choices of all banks strategic complements (Farhi and Tirole, 2017). The subsidy reinforces the already strong precautionary lending motive - banks choose higher equilibrium leverage because it allows them to accumulate more net worth with less downside risk.

Figure 20 presents the macroeconomic effects of the TBTF problem. As usual, we are considering two regimes with an instantaneous transition. We see that the TBTF hazard reduces aggregate output, net output, and financial intermediation activity. The subsidy directly affects big banks who already have satiated demand, thus the negative net effect on aggregate quantities. The banking sector has also become more concentrated. The positive impact on aggregate default risk is the result of strategic complementarity in risk-taking - the economy is less efficient and less stable. On the other hand, the aggregate markup and the risk premium decline mildly. The mechanism of the trilemma goes through.

Figure 20: **Macroeconomic Effects of TBTF**



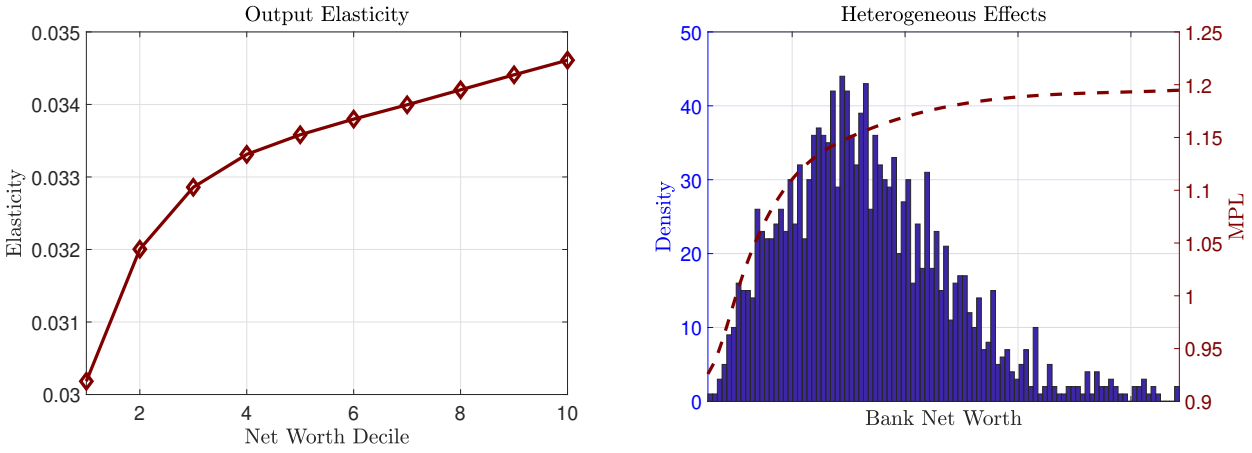
Notes: macroeconomic effects of switching on the TBTF subsidy.

5.4 Targeted Stabilization Policies

Equity Injections Credit policy has been modelled in several representative-agent Macroeconomic frameworks, for example Curdia and Woodford (2010, 2016), He and Krishnamurthy (2013). We move beyond systematic credit policy analysis and estimate conditional macro elasticities when equity injections are allowed on any *individual* bank in the distribution. We proceed in two steps. First, we break the distribution of bank net worth into ten bins (deciles). For each decile $\iota = 1 \ldots 10$, we assume that the government increases by one percent the net worth of each bank in ι but not anywhere else in the distribution. Second, we compute the macro elasticity with respect to targeted policies using the same, equilibrium MPL distribution but integrating it over different ex-post distributions of bank net worth after the equity injections took place. We thus run ten separate experiments, one per each decile of the size distribution, and compute the conditional impact on aggregate output ten separate times.

Figure 21 plots the result. We observe that there are efficiency gains from injecting equity into large intermediaries. The elasticity of aggregate output with respect to decile-specific credit policies is an upward-sloping line. This result is driven by the shape of the MPL distribution - larger banks have a greater equilibrium MPL, which is in turn due to big banks having lower marginal costs and relative prices. Abstracting from any normative implications or second-level

Figure 21: **Macroeconomic Effects of Targeted Equity Injections**



Notes: Responses of aggregate output to targeted, decile-specific bank equity injections.

effects on financial stability or systemic risk, if the objective of the government is purely to stimulate aggregate lending and demand, then "bailing out" big banks yields a bigger bang for the buck.[13]

Liquidity Facility Financial crises are typically associated with tightening of liquidity constraints. As opposed to the lack of credit worthiness of borrowers, it is the lack of liquidity on the credit supply side that contributes to rising excess returns. In our model, banks face a liquidity constraint in the form of the moral hazard-induced cap on leverage-taking. The fraction of divertible assets - λ - controls the degree of constraint tightness and is generally part of the exogenous environment. We now suppose that the government can step in and augment λ on behalf of private lenders. In particular, we allow λ to be relaxed on *any* bank in the distribution. In practice, this intervention can be mapped to discount window lending to banks secured by the credit portfolio.

In order to facilitate the cleanest possible analysis, we assume that the leverage constraint binds on all banks in the distribution.[14] With the binding leverage constraint, it is straightforward to solve for the bank-specific leverage ratio:

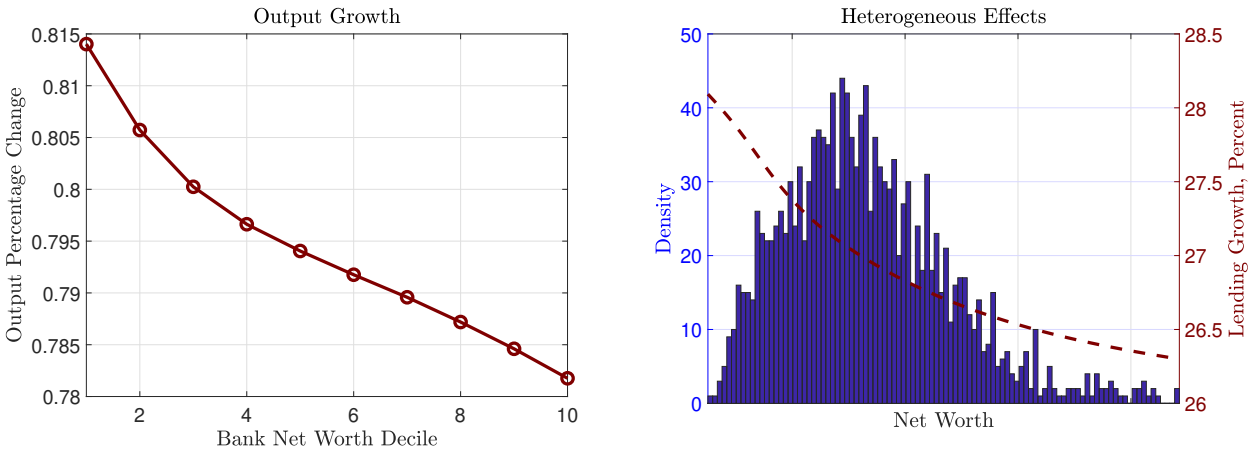
$$\phi(j) = \frac{\nu_a(j)}{\lambda - \mu_a(j)} \tag{34}$$

where, as before, $\phi(j)$ is market leverage, $\nu_a(j)$ is the discounted cost of bank liabilities, $\mu_a(j)$ are excess returns on the risky asset. Notice how according to this formula, relaxation of liquidity conditions (as proxied by a reduction in λ) increases banks appetite for leverage. Everything else equal, this raises credit supply in the market.

[13] 'These bailouts are unexpected and do not generate additional moral hazard frictions ex-ante. The implicit bailout subsidy is internalized in the previous Section 5.3.

[14] This is a realistic assumption given that these types of policies are usually only implemented in crisis episodes, precisely when liquidity and leverage constraints of market lenders tighten.

Figure 22: **Macroeconomic Effects of Targeted Liquidity Facilities**



Notes: Responses of aggregate output to targeted, decile-specific liquidity facilities.

We proceed by assuming that the government intervenes by lowering λ_ι on decile $\iota = 1 \ldots 10$ of the banks net worth distribution by 10% relative to the baseline value of 0.1. The exogenous shock is thus invariant to the region of the distribution which is targeted. The only variant in this policy intervention is the decile of the bank net worth distribution. For each of the ten policy counterfactuals, we compute the conditional output elasticity.

Figure 22 presents the result. We see that the differential impact of this policy is concentrated in the left tail of the distribution - smaller banks increase their credit by more. On the left panel we see how this translates into a downward-sloped output elasticity curve. This result arises because the marginal effect of λ_ι on ϕ_ι is negative and declining with bank size due to diminishing marginal costs of funds. Because we assumed that the constraint is always binding, we have therefore isolated the intensive margin of the total effect. Suppose now that the constraint can bind occasionally. The Lagrange multiplier on the constraint declines in bank size. In other words, the constraint is generally slack for big banks and binding for small banks. Relaxation of the moral hazard friction is therefore much more likely to differentially benefit small banks even if we allow the constraint to bind occasionally. Results of Figure 22 would therefore not change.

5.5 Intermediary Asset Pricing

He et al. (2016) and Adrian et al. (2014), among others, have popularized the intermediary asset pricing view: in contrast to conventional models, the true pricing kernel is a function of intermediary balance sheet ratios such as capital or leverage. This literature, however, relies on the representative agent assumption and abstracts from distributional dimensions. One exception is Ma (2018) who argues that intermediaries that have low levels of net worth and face tighter financing

Table 4: **Asset Pricing Moments**

	Risk-Free Rate	Risky Return	Risk Premium
No Banks	1.004	1.004	0
Homogenous Bank	1.004	1.0265	0.0225
Only Monopolistic Competition	1.004	1.0274	0.0234
Only Idiosyncratic Risk	1.0028	1.0313	0.0285
Baseline	1.0046	1.0448	0.0402

Notes: main asset pricing moments for various versions of the model.

constraints are the true marginal investors in a large class of intermediary asset pricing models.

The banker's Euler equation can be re-formulated into a classic asset pricing formula for the risk premium:

$$\mathbb{E}_t\left[R^T_{t+1}(j) - R^{rf}_t(j)\right] = \underbrace{\frac{\lambda \overbrace{\varphi(j)}^{\text{Lagrange Multiplier}}}{\mathbb{E}_t\left(\hat{\Lambda}_{t+1}(j)\right)}}_{\text{Liquidity Premium}} + \underbrace{\nu(j)}_{\text{Default Premium}} \underbrace{+ \operatorname{cov}\left[-\frac{\hat{\Lambda}_{t+1}(j)}{\mathbb{E}_t\left(\hat{\Lambda}_{t+1}(j)\right)}, R^T_{t+1}(j)\right]}_{\text{Risk Premium}} \quad \forall j$$

Where $\varphi(j)$ is the Lagrange multiplier on the moral hazard (leverage) constraint. Note that the equation must hold for every bank (j) in the distribution. If financial frictions are switched off, then the intermediation spread is zero. Excess returns in the baseline economy arise for two reasons. First, if the hard leverage constraint binds for any given bank j, or has a positive probability of binding in the future, then external funds are harder to obtain. This is the liquidity-induced external finance premium. Second, presence of bank default risk requires additional ex-ante compensation. Note that the canonical risk premium is absent in the stationary equilibrium because we abstract from aggregate uncertainty.

Table 4 presents key asset pricing moments from the framework under different assumptions. Without any heterogeneity, the liquidity and default risk channels generate a premium of 2.25%. Adding monopolistic competition with variable markups and uninsurable idiosyncratic shocks gets us a risk premium of 4.02%. This occurs because both liquidity and default risk premia are concentrated in the left tail of the distribution. Heterogeneity switches on the extensive margin, and a large equilibrium share of low-net-worth banks raises aggregate riskiness of the economy. Without aggregate uncertainty and relying solely on idiosyncratic shocks and the model structure, we therefore can explain roughly 2/3 of the aggregate unconditional risk premium.

6 Conclusion

In this article I develop a novel tractable, dynamic general equilibrium framework for positive and normative analysis of macroeconomic transmission through bank heterogeneity. The model introduces two workhorse approaches in modern macro-finance - uninsurable idiosyncratic risk and imperfect competition - into the banking sector of the seminal Gertler and Kiyotaki (2010) environment. The calibrated model delivers disperse, concentrated cross-sectional distributions of bank assets, net worth, leverage, markups, marginal costs, relative prices, default risk, and deposit rates. The model is validated by replicating key empirical patterns: cross-sectional relations between bank size, book leverage, exit risk, and markups. Micro-banking heterogeneity and macroeconomic fluctuations are linked through *Marginal Propensity to Lend* heterogeneity, which determines the elasticity of aggregate demand to changes in bank net worth.

Bank heterogeneity matters for the design of various economic policies that run through the bank lending and market power channels. I analyze different regulatory schemes such as deposit guarantees and heterogeneous capital requirements. I also study optimal policy in a fully constrained-efficient version of the economy. Policy analysis points at a *trilemma* for bank regulation. There is a trilateral trade-off between financial competition, stability, and efficiency. Any policy intervention that attempts to improve one of these three factors, necessarily worsens at least one of the remaining two. Through the lenses of this trilemma, I characterize predictions and implications of the framework for the rise of banking concentration, emergence of fintech-intermediated credit, the workings of the too-big-to-fail externality, unconventional fiscal and monetary policy interventions, and intermediary asset pricing.

My model is tractable and can be readily extended to include additional parts. First, an open-economy extension could be introduced, allowing us to study endogenous global financial cycles that are driven by heterogeneous, imperfectly competitive intermediaries. Second, the model in its present form abstracts from aggregate risk. Jamilov and Monacelli (2020) build on my framework and introduce aggregate uncertainty and a dynamic distribution of bank size. They study novel channels of business cycle amplification that arise from dynamic bank heterogeneity. Finally, the current model has no role for conventional monetary policy. An extension with nominal rigidities could uncover a powerful channel of transmission that runs through bank heterogeneity in risk-taking and market power.

References

Adrian, T., E. Etula, and T. Muir (2014): "Financial Intermediaries and the Cross-Section of Asset Returns," *Journal of Finance*, 69(6), 2557–2596.

Adrian, T. and H. S. Shin (2010): "Liquidity and Leverage," *Journal of Financial Intermediation*, 19(3), 418–437.

——— (2014): "Procyclical Leverage and Value-at-Risk," *Review of Financial Studies*, 27(2), 373–403.

Aiyagari, R. (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *Quarterly Journal of Economics*, 109(3), 659–684.

Anderson, S., A. D. Palma, and J. Thisse (1989): "Demand for Differentiated Products, Discrete Choice Models, and the Characteristics Approach," *The Review of Economic Studies*, 56(1).

Atkeson, A. G., A. d'Avernas, A. L. Eisfeldt, and P.-O. Weill (2018): "Government Guarantees and the Valuation of American Banks," .

Auclert, A. (2019): "Monetary Policy and the Redistribution Channel," *American Economic Review*, 109.

Beck, T., A. Demirguc-Kunt, and R. Levine (2006): "Bank concentration, competition, and crises: First results," *Journal of Banking Finance*, 30, 1581–1603.

Begenau, J., S. Bigio, J. Majerovitz, and M. Vieyra (2020): "A Q-Theory of Banks," *Manuscript*.

Begenau, J. and T. Landvoigt (2020): "Financial Regulation in a Quantitative Model of the Modern Banking System," *Working Paper*.

Benetton, M. (2021): "Leverage Regulation and Market Structure: A Structural Model of the UK Mortgage Market," *Journal of Finance*.

Benhabib, B., A. Bisin, and M. Luo (2019): "Wealth distribution and social mobility in the US: A quantitative approach," *American Economic Review*, 109.

Berger, A. N. and T. H. Hannan (1998): "The Efficiency Cost of Market Power in the Banking Industry: A Test of the "Quiet Life" and Related Hypotheses," *The Review of Economics and Statistics*, 80.

Berger, A. N. and L. J. Mester (1997): "Inside the black box: What explains differences in the efficiencies of financial institutions?" *Journal of Banking Finance*, 21.

Bewley, T. (1977): "The permanent income hypothesis: A theoretical formulation," *Journal of Economic Theory*, 16, 252 – 292.

Bianchi, J. and S. Bigio (2020): "Banks, Liquidity Management and Monetary Policy," *Manuscript*.

Bigio, S. and Y. Sannikov (2021): "A Model of Credit, Money, Interest, and Prices," .

Blanchard, O. and N. Kiyotaki (1987): "Monopolistic Competition and the Effects of Aggregate Demand," *American Economic Review*, 77(4).

Bocola, L. (2016): "The Pass-Through of Sovereign Risk," *Journal of Political Economy*, 124 (4).

Boissay, F., F. Collard, and F. Smets (2016): "Booms and Banking Crises," *Journal of Political Economy*, 124(2).

Bordalo, P., N. Gennaioli, and A. Shleifer (2018): "Diagnostic Expectations and Credit Cycles," *The Journal of Finance*, 73(1), 199–227.

Boyd, J. and G. D. Nicolo (2005): "The Theory of Bank Risk Taking and Competition Revisited," *Journal of Finance*, 60(3).

Brunnermeier, M. and L. Pedersen (2009): "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 22, 2201–2238.

Brunnermeier, M. and Y. Sannikov (2014): "A Macroeconomic Model with a Financial Sector," *American Economic Review*, 104(2), 379–421.

Capelle, D. (2019): "Competition vs. Stability: Oligopolistic Banking System with Run Risk," *Working Paper*.

Christiano, L. and D. Ikeda (2013): "Leverage Restrictions in a Business Cycle Model," *NBER Working Paper 18688*.

Claessens, S., J. Frost, G. Turner, and F. Zhu (2018): "Fintech credit markets around the world: size, drivers and policy issues," *BIS Quarterly Review*.

Coimbra, N. and H. Rey (2019): "Financial Cycles with Heterogeneous Intermediaries," *NBER Working Paper*, 23245.

Constancio, V. (2016): "Challenges for the European Banking Industry," *Conference on "European Banking Industry: what's next?"*.

Corbae, D. and P. D'Erasmo (2020a): "Capital Requirements in a Quantitative Model of Banking Industry Dynamics," *NBER Working Paper*, 25424.

——— (2020b): "Rising bank concentration," *Journal of Economic Dynamics and Control*, 115.

Corbae, D. and R. Levine (2018): "Competition, Stability, and Efficiency in Financial Markets," *Jackson Hole Symposium: Changing market Structure and Implications for Monetary Policy*.

Curdia, V. and M. Woodford (2010): "Credit Spreads and Monetary Policy," *Journal of Money, Credit and Banking*, 42.

——— (2016): "Credit Frictions and Optimal Monetary Policy," *Journal of Monetary Economics*, 84.

Davydiuk, T. (2019): "Dynamic Bank Capital Requirements," *Journal of Financial Economics*, Working Paper.

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

Dempsey, K. (2020): "Capital Requirements with Non-Bank Finance," *Working Paper*.

Diamond, D. (1984): "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies*, 51(3).

Dixit, A. and J. Stiglitz (1977): "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67(3).

Drechsler, I., A. Savov, and P. Schnabl (2017): "The deposits channel of monetary policy," *Quarterly Journal of Economics*, 132, 1819–1876.

Edmond, C., V. Midrigan, and D. Xu (2018): "How Costly are Markups," *NBER Working Paper*.

Egan, M., A. Hortacsu, and G. Matvos (2017): "Deposit Competition and Financial Fragility: Evidence from the US Banking Sector," *American Economic Review*, 107(1).

Farhi, E. and J. Tirole (2017): "Shadow Banking and the Four Pillars of Traditional Financial Intermediation," *NBER Working Paper*, 23930.

Fostel, A. and J. Geanakoplos (2008): "Leverage Cycles and the Anxious Economy," *American Economic Review*, 98.

Gabaix, X. (2009): "Power Laws in Economics and Finance," *Annual Review of Economics*, 1.

——— (2011): "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 79(3).

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016): "The Dynamics of Inequality," *Econometrica*, 84.

GALAASEN, S., R. JAMILOV, R. JUELSRUD, AND H. REY (2020): "Granular Credit Risk," *NBER Working Paper 27994*.

GERTLER, M. AND P. KARADI (2011): "A Model of Unconventional Monetary Policy," *Journal of Monetary Economics*, 58(1), 17–34.

GERTLER, M. AND N. KIYOTAKI (2010): "Financial Intermediation and Credit Policy in Business Cycle Analysis," *Handbook of Monetary Economics*, 3, 547–599.

GERTLER, M., N. KIYOTAKI, AND A. PRESTIPINO (2016): "Wholesale Banking and Bank Runs in Macroeconomic Modelling of Financial Crises," *Handbook of Macroeconomics*, 2.

——— (2020): "A Macroeconomic Model with Financial Panics," *Review of Economic Studies*, 87(1).

GOLDSTEIN, I., A. KOPYTOV, L. SHEN, AND H. XIANG (2020): "Bank Heterogeneity and Financial Stability," *NBER Working Paper 27376*.

GORTON, G. AND A. METRICK (2010): "Regulating the Shadow Banking System," *Brookings Papers on Economic Activity*, Fall.

GRENVILLE, S. (2016): "Systemic Risk, Crises, and Macroprudential Regulation, by Xavier Freixas, Luc Laeven and JosÃ©-Luis PeydrÃ^3(MITPress, Cambridge, MA, 2015), pp.xii + 472.″*Economic Record*, 92, 313 – –314.

GROMB, D. AND D. VAYANOS (2002): "Equilibrium and welfare in markets with financially constrained arbitrageurs," *Journal of Financial Economics*, 66.

HALDANE, A. (2010): "The $100 Billion Question. Commentary," *Bank of England*.

HE, Z., B. KELLY, AND A. MANELA (2016): "Intermediary Asset Pricing: New Evidence from Many Asset Classes," *Journal of Financial Economics*, Forthcoming.

HE, Z. AND A. KRISHNAMURTHY (2013): "Intermediary Asset Pricing," *Journal of Financial Economics*, 103(2), 732–770.

HELLMAN, T., K. MURDOCK, AND J. STIGLITZ (2000): "Liberalization, Moral Hazard in Banking, and Prudential Regulation: Are Capital Requirements Enough?" *American Economic Review*, 90(1).

HUBERMAN, G. (2015): "Familiarity Breeds Investment," *The Review of Financial Studies*, 14, 659–680.

HUGGETT, M. (1990): "The Risk-Free Rate in Heterogeneous Agent Economies," *Manuscript, University of Minnesota*.

IMROHOGLU, A. (1996): "Costs of Business Cycles with Indivisibilities and Liquidity Constraints," *Journal of Political Economy*, 1364–83.

IVASHINA, V. (2009): "Asymmetric information effects on loan spreads," *Journal of Financial Economics*, 92(2).

JAMILOV, R. (2020): "Credit Market Power: Branch-level Evidence from the Great Finanial Crisis," *Working Paper*.

JAMILOV, R. AND T. MONACELLI (2020): "Bewley Banks," *CEPR Discussion Paper 15428*.

JERMANN, U. AND V. QUADRINI (2013): "Macroeconomic Effects of Financial Shocks," *American Economic Review*, 102(1), 238–271.

JUELSRUD, R. E. AND E. G. WOLD (2020): "Risk-weighted capital requirements and portfolio rebalancing," *Journal of Financial Intermediation*, 41, 100806.

KAPLAN, G., B. MOLL, AND G. VIOLANTE (2018): "Monetary Policy According to HANK," *American Economic Review*, 108(3).

KEELEY, M. C. (1990): "Deposit Insurance, Risk, and Market Power in Banking," *The American*

Economic Review, 80.

KIMBALL, M. (1995): "The quantitative analytics of the basic neomonetarist model," *Journal of Money, Credit and Banking*, 27(4).

KLENOW, P. J. AND J. L. WILLIS (2016): "Real Rigidities and Nominal Price Changes," *Economica*, 83.

KORINEK, A. AND M. NOWAK (2016): "Risk-Taking Dynamics and Financial Stability," *Manuscript*.

LAEVEN, L., R. LEVINE, AND S. MICHALOPOULOS (2015): "Financial innovation and endogenous growth," *Journal of Financial Intermediation*, 24.

LAEVEN, L. AND F. VALENCIA (2018): "Systemic Banking Crises Database: An Update," *IMF Working Paper*, 18/208).

LEE, S., R. LUETTICKE, AND M. RAVN (2020): "Financial Frictions: Macro vs Micro Volatility," *CEPR DP*, 15133.

MA, S. (2018): "Heterogeneous Intermediaries and Asset Prices," *Working Paper*.

MALIAR, L., S. MALIAR, AND F. VALLI (2010): "Solving the incomplete markets model with aggregate uncertainty using the Krusellâ€"Smith algorithm," *Journal of Economic Dynamics and Control*, 34.

MARTINEZ-MIERA, D. AND R. REPULLO (2010): "Does Competition Reduce the Risk of Bank Failure?" *The Review of Financial Studies*, 23.

MCFADDEN, D. (1984): "Econometric Analysis of Qualitative Response Models," *Grilliches, Z. and Intriligator, M. (eds) Handbook of Econometrics*, 2.

MEHRA, R., F. PIGUILLEM, AND E. C. PRESCOTT (2011): "Costly financial intermediation in neoclassical growth theory," *Quantitative Economics*, 2.

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

MIDRIGAN, V., C. EDMOND, AND D. XU (2018): "How Costly are Markups," *NBER Working Paper*, 24800.

NGUYEN, T. (2015): "Bank Capital Requirements: A Quantitative Analysis," *Working Paper*.

NUNO, G. AND C. THOMAS (2016): "Bank Leverage Cycles," *American Economic Journal: Macroeconomics*, Forthcoming.

PASQUALINI, A. (2021): "Markups, Markdowns and Bankruptcy in the Banking Industry," *Working Paper*.

REINHART, C. M. AND K. S. ROGOFF (2009): "The Aftermath of Financial Crises," *American Economic Review*, 99.

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

RIOS RULL, V., T. TAKAMURA, AND Y. TERAJIMA (2020): "Banking Dynamics, Market Discipline and Capital Regulations," *Manuscript*.

ROMER, C. D. AND D. H. ROMER (2017): "New Evidence on the Aftermath of Financial Crises in Advanced Countries," *American Economic Review*, 107.

SCHULARICK, M. AND A. M. TAYLOR (2012): "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870-2008," *American Economic Review*, 102.

STAVRAKEVA, V. (2019): "Monetary Policy Spill-Overs in Emerging Markets and Imperfect Banking Sector Competition," *Manuscript*.

STERN, G. AND R. FELDMAN (2004): "Too Big to Fail: The Hazards of Bank Bailout," *Brookings Institution Press*.

SUFI, A. (2007): "Bank Lines of Credit in Corporate Finance: An Empirical Analysis," *Review of Financial Studies*, 22.

VAN NIEUWERBURGH, S. AND L. VELDKAMP (2009): "Information immobility and the home bias puzzle," *The Journal of Finance*, 64, 1187–1215.

WHEELOCK, D. C. AND P. WILSON (2012): "Do Large Banks Have Lower Costs? New Estimates of Returns to Scale for U.S. Banks," *Journal of Money, Credit and Banking*, 44.

WHEELOCK, D. C. AND P. W. WILSON (2018): "The evolution of scale economies in US banking," *Journal of Applied Econometrics*, 33.

A Proofs

A.1 Bank Markup and Marginal Cost Decomposition

Assumptions: Bank-level choices are made while $\bar{R}(j)$, $R^T(j)$, $\nu(j)$ are taken as given. Leverage constraint is slack. Without loss of generality, assume $\zeta_1 = \zeta_2$.

Show that the bank price-setting rule is:

$$\frac{p(j)}{P} = \mu(x)\frac{k(j)^{\zeta_2 - 1}}{R^T(j) - \bar{R}(j)}$$

Each bank j solves

$$\max_{k(j)} \left\{ \tilde{\Lambda}\Big(1 - \upsilon(j)\Big)\left[R^T(j)p(j)k(j) - \bar{R}\Big(p(j)k(j) - n(j)\Big) - \frac{1}{\zeta_2}k(j)^{\zeta_2}\right]\right\} \quad \text{s.t.} \quad p_t(j) = \Upsilon'\left(\frac{k(j)}{K}\right)Z_t$$

where $Z \coloneqq \left(\int_0^1 \Upsilon'\left(\frac{k(j)}{K}\right)\frac{k(j)}{K}dj\right)^{-1}$. The first order condition is

$$\tilde{\Lambda}\Big(1 - \nu(j)\Big)\left\{\Big(R^T(j) - \bar{R}(j)\Big)\left(p(j) + k(j)\frac{\partial p(j)}{\partial k(j)} - k(j)^{\zeta_2 - 1}\right)\right\} = 0$$

Assume that the impact of p(j) on the aggregate index P is not internalized. The elasticity is:

$$\frac{\partial k(j)}{\partial p(j)}\frac{p(j)}{k(j)} = x^{-\frac{\epsilon}{\theta}}$$

where x is relative bank size. The markup function $\mu(x)$ is:

$$\mu(x) = \frac{\theta x^{-\frac{\epsilon}{\theta}}}{\theta x^{-\frac{\epsilon}{\theta}} - 1}$$

The marginal cost MC(j) is given by:

$$MC(j) \coloneqq \frac{1}{R^T(j) - \bar{R}(j)}k(j)^{\zeta_2 - 1}$$

The price-setting rule given marginal costs is thus:

$$\frac{p(j)}{P} = \frac{\theta x^{-\frac{\epsilon}{\theta}}}{\theta x^{-\frac{\epsilon}{\theta}} - 1}\frac{k(j)^{\zeta_2 - 1}}{R^T(j) - \bar{R}(j)}$$

where the first term on the right hand side is the markup and the second term is the marginal cost.

Constant Markup Whenever $\epsilon = 0$ the relative price price rule becomes:

$$p(j) = \frac{\theta}{\theta - 1}MC(j)$$

where $\frac{\theta}{\theta-1}$ is the constant markup over the marginal cost which is now:

$$MC(j) := \frac{1}{R^T(j) - \bar{R}(j)} \left[\left(\frac{p(j)}{P} \right)^{-\theta} K \right]^{\zeta_2 - 1}$$

Solving out aggregate prices gives:

$$\frac{p(j)}{P} = \left[\frac{\theta}{\theta - 1} \frac{1}{R^T(j) - \bar{R}(j)} \frac{1}{P} K^{\zeta_2 - 1} \right]^{\frac{1}{1+\theta(\zeta_2 - 1)}}$$

Note that this equation resembles the canonical price rule in Blanchard and Kiyotaki (1987).
□

A.2 Bank Scale Variance

Guess that the solution to the dynamic problem is a value function $V\Big(n(j),\xi(j)\Big) = \vartheta\Big(n(j),\xi(j)\Big)n(j)$. Define the default risk-adjusted stochastic discount factor $\tilde{\Lambda} = \Big(1-\upsilon(j)\Big)\Lambda\Big(1-\sigma+\sigma\vartheta\Big(n(j),\xi(j)\Big)\Big)$. The solution to the program is a system of equations:

$$\mathbb{E}\left[\tilde{\Lambda}\left(R^{T}(j)-\bar{R}(j)\right)\right] = \lambda\varphi\Big(n(j),\xi(j)\Big)$$

$$\varphi\Big(n(j),\xi(j)\Big)\left[\vartheta\Big(n(j),\xi(j)\Big)-\lambda\phi(j)\right] = 0$$

Substituting the optimality conditions together with the guess into the objective function gives

$$\vartheta\Big(n(j),\xi(j)\Big) = \varphi\Big(n(j),\xi(j)\Big)\vartheta\Big(n(j),\xi(j)\Big) + \mathbb{E}\Big[\tilde{\Lambda}\Big(\bar{R}(j)-\frac{\frac{1}{\zeta_1}k(j)^{\zeta_2}}{n(j)}\Big)\Big]$$

Solving for $\vartheta\Big(n(j),\xi(j)\Big)$ yields

$$\vartheta\Big(n(j),\xi(j)\Big) = \frac{\mathbb{E}\Big[\tilde{\Lambda}\Big(\bar{R}(j)-\frac{\frac{1}{\zeta_1}k(j)^{\zeta_2}}{n(j)}\Big)\Big]}{1-\varphi\Big(n(j),\xi(j)\Big)}$$

And the Lagrange multiplier on the leverage constraint is

$$\varphi\Big(n(j),\xi(j)\Big) = \max\left[1-\frac{\mathbb{E}\Big[\tilde{\Lambda}\Big(\bar{R}(j)-\frac{\frac{1}{\zeta_1}k(j)^{\zeta_2}}{n(j)}\Big)\Big]}{\lambda\phi(j)},0\right]$$

Note that when ϵ = then market leverage becomes $\phi(j) = k(j)^{\frac{\theta-1}{\theta}}K^{\frac{1}{\theta}}Pn(j)^{-1}$. The guess is verified if $\varphi\Big(n(j),\xi(j)\Big) < 1$. Net worth-dependency is guaranteed by $\zeta_2 \neq 1$ (for a given $\zeta_1 \neq 0$) so that each bank with a different n(j) chooses its own leverage ratio $\phi(j)$. Furthermore, with $\kappa > 0$, $\phi(j)$ also explicitly depends on $\xi(j)$.[15] □

[15]Note that $\{\epsilon,\theta\}$ do not impact scale-dependency but do change the level and curvature of the $\vartheta\Big(n(j),\xi(j)\Big)$ surface.

Table 5: **Description of Banking Variables**

Variable Name	Description	Source
Assets	Total assets (RCFD2170)	Call reports
Equity	Total assets (RCFD2170) - total liabilities (RCFD2948)	Call reports
Leverage ratio	Assets / equity	Call reports
Capital ratio	Equity / assets	Call reports
Deposit expenses	Interest expense on domestic deposits. Equals total interest expense on deposits (RIAD4170) - interest expense on foreign deposits (RIAD4172)	Call reports
Non-interest expenses	Total noninterest expenses (RIAD4093)	Call reports
Net interest margin	Net interest income (RIAD4074) / assets	Call reports

Notes: This table details the construction, definition, and sourcing of the empirical variables used for model parameterization.

ONLINE APPENDIX

"A Macroeconomic Model with Heterogeneous Banks"

by Rustam Jamilov

B Microfoundations and Extensions

B.1 CES Monopolistic Financial Intermediation

This section provides a brief theoretical foundation for the representative-agent capital goods producer's CES credit demand system. My approach follows closely McFadden (1984) and Anderson et al. (1989). We focus on the analytically more convenient case when $\epsilon = 0$. Assume there are M borrowers and H banks. Each banker i posts its price schedule. Each borrower j observes the price menu and receives an idiosyncratic preference shock ϵ_{ij} which is borrower-creditor specific.

Assume the production function of a borrowing firm j is $\log k(j)$. All borrowers are indexed by their favorite bank branch $\bar{\epsilon}$. They suffer disutility measured in Euclidean distance between their preferred type and any given type i. Unit cost of that disutility, as well as the distance between varieties have been set to unity. Profit function of each firm takes on the following form.

$$Q_i(\bar{\epsilon}; k_i) = \underbrace{\log k_i + Y - p_i k_i}_{\text{Homogenous across j}} - \underbrace{\sum_{k=1}^{M} \left(\bar{\epsilon}^k - \epsilon_i^k\right)^2}_{\text{Heterogeneous across j}} \qquad i=1 \dots H \tag{35}$$

The first term in the equation is common across all borrowers and is bank-specific. The second term is the bank-borrower fixed effect that captures disutility from not borrowing from the ideal branch $\bar{\epsilon}$. Without loss of generality, we impose $M = H - 1$ for analytical convenience. We define *credit market access* as the set of consumers that are indifferent between borrowing from any two branches n:

$$\bar{E}^j = \frac{\log \frac{p(j)}{p_n}}{4} \tag{36}$$

The choice variables are (a) which branch to borrow from and (b) how much k(j). The price of the loan p(j) corresponds to the price on a claim on returns to capital in the main text. Y is endogenous real income that in equilibrium will equal K, i.e. the book value of capital after assembly and aggregation.

Every borrower in the credit market access space borrows $\frac{1}{p_n}$ units of differentiated loans from bank n. Demand k_n becomes:

$$k_n = \frac{1}{p_n} \int_{-\infty}^{\bar{\epsilon}_1} \dots \int_{-\infty}^{\bar{\epsilon}_{n-1}} f\left(\bar{E}^j\right) d\bar{\epsilon} \tag{37}$$

Where we assume that k_n is strictly positive for all prices p(j), is n-1 times continuously differentiable, and all cross-price derivatives are positive for all i and j. Solution for the credit demand function above involves taking n-1 derivatives of k_n w.r.t. $p_1, \dots, p_{H-1}$:

$$\frac{\partial^{H-1} k_n}{\partial p_1 \dots \partial p_{H-1}} = \frac{1}{p_1 \cdots p_n} 4^{1-H} f\left(\bar{E}^j\right) \tag{38}$$

We assume that the firm borrower demand function is logistic in the cross-price differential p(j)−p(i) for any two branches i and j. The density function associated with a logit credit demand is given

by:

$$f(\bar{\epsilon}) = H\frac{4}{\bar{\theta}}^{H-1}(H-1)!\frac{\prod_{i=1}^{H-1}\exp\left(-4/\bar{\theta}\epsilon^{i}\right)}{\left[1+\sum(j)^{H-1}\exp\left(-4/\bar{\theta}\epsilon^{j}\right)\right]^{H}} \tag{39}$$

Plugging our model-specific credit market access variable into the logit density, and evaluating the first order condition yields

$$\frac{\partial^{H-1}k_n}{\partial p_1 \ldots \partial p_{H-1}} = H\bar{\theta}^{1-H}(H-1)!\frac{\prod(j)^{H}p(j)^{-1/\bar{\theta}-1}}{\left[\sum(j)^{H}p(j)^{-1/\bar{\theta}}\right]^{H}} \tag{40}$$

Integrating gives us optimal credit demand

$$k_n = Hp_n^{-1/\bar{\theta}-1}\left[\sum(j)^{H}p(j)^{-1/\bar{\theta}}\right]^{-1} \tag{41}$$

Now, we impose the following parameter restriction: $\bar{\theta} = \frac{1}{\theta-1}$. Furthermore, impose the accounting identity that the total sum of firm-level loans is equal to the income of the representative capital goods producer: $Hk(j) = K$. We retrieve the CES credit demand function of firm j in main text:

$$k(j) = \left(\frac{p(j)}{P}\right)^{-\theta}K \tag{42}$$

We have thus shown that the representative-agent capital goods producer setup in main text is isomorphic to a heterogeneous-borrower environment with idiosyncratic preferences for branch amenities. The logit parameter $\bar{\theta}$ captures the variance of borrower preferences and maps conveniently to the CES elasticity θ. The relationship is inversed, so a higher $\bar{\theta}$ is associated with a lower elasticity of credit supply, i.e. greater credit market power. In the limit, if $\bar{\theta} \to \infty$ we recover a case with a single pure monopoly provider of credit. As $\bar{\theta} \to 0$ we recover the case of perfect competition in the banking sector. Because the problem discussed in this section is static, and assuming the distribution of shocks is time-invariant, heterogeneous firms would solve the same static problem every period and arrive at the same solution. It's therefore convenient, as we do in the main text, to work the representative-agent representation of this distribution.

B.2 Portfolio Returns

In this section we explain how our formulation of total portfolio returns (Equation 7) can be microfounded. Suppose there are N banks and credit markets. These credit markets could be understood in at least three different ways: units in geographical space (counties), segmented industries, or segmented financial varieties (products/services). The model is isomorphic to any of these interpretations. Now suppose that each bank b specializes in one credit market c and overweighs it by $0 < \kappa < 1$. Concentration can be motivated by a variety of theories, including but not limited to "home" bias in bank lending (Juelsrud and Wold, 2020), asymmetric information (Sufi, 2007; Ivashina, 2009; Van Nieuwerburgh and Veldkamp, 2009), or behavioral biases (Huberman, 2015; Bordalo et al., 2018). Assume that market-specific returns R^j are not diversifiable/insurable. This assumption can be motivated by the empirical findings in Galaasen et al. (2020). Then, the

bank-specific portfolio return can be written as:

$$R^b = \sum_j^N \frac{1}{N} R^j + \kappa R^c - \kappa R^{-c} \tag{43}$$

where R^{-c} is the return on a portfolio that excludes the bank's favorite market c. Now, we assume that N is large enough such that R^{-c} is approximately equal to the return on the market portfolio R^k. That is, credit markets are atomistic:

$$R^b \approx R^k + \kappa R^c - \kappa R^k = \kappa R^c + (1-\kappa) R^k$$

Which is the same formulation that we used in Equation 7, except that in the model R^c is $\xi(j)$ and follows an autoregressive process. Now, total return across all banks can be written as:

$$R^{\text{total}} = \sum_b^N \frac{1}{N} \kappa R^c + (1-\kappa) \sum_b^N R^k = R^k$$

That is, in the aggregate, credit market-specific idiosyncratic returns vanish and banks are exposed only to the systematic component of returns R^k. What makes idiosyncratic return risk an intertemporal problem for banks are (a) scale variance and (b) persistence of $\xi(j)$.

B.3 Two-Sector Extension

The baseline economy in the main text features a single capital goods sector which is intermediated by imperfectly competitive banks. It's possible to generalize our setup to two types of capital goods. Suppose the first capital good K_{at} is imperfectly differentiated across the mass of banks H_t. These are the financial varieties which we discuss in main text. The second good type K_{bt} is a perfect substitute across lenders. This proxies standard fixed-term commercial loans which are homogenized across banks, who in turn face perfect competition in this market. We continue to assume that there is a representative capital goods producer that is financially constrained and requires bank funds in order to produce the capital stock. The production stage of the capital stock now consists of two steps. First, we determine the equilibrium fraction of differentiated capital goods K_{at}. The capital goods firm solves the following problem:

$$\min_{K_{at}, K_{bt}} P_t K_{at} + K_{bt} \quad \text{s.t.} \quad K_{at}^{\chi} K_{bt}^{1-\chi} = K_t \tag{44}$$

Where $0 < \chi < 1$ is the elasticity of substitution across the types of capital goods. The solution delivers a set of two familiar equations: $P_t K_{at} = \chi K_t$ and $K_{bt} = (1-\chi) K_t$. That is, the share of financial varieties in the economy is time-invariant and is equal to χ. The second stage of the problem is determination of the demand for individual varieties $k_t(j)$ within the K_{at} sector.

The parameter χ could be taken to the data and mapped to the scale and intensity of shadow banking activities before the Crisis (Gorton and Metrick, 2010). Parameter statics in χ could be used to simulate advancements in financial innovation and/or the rise of complexity in the credit market (Grenville, 2016; Laeven et al., 2015).

C Alternative Stabilization Policies

C.1 Direct Lending Facility

In this section we study an alternative form of targeted credit policies - a direct lending facility. Broadly speaking, we are referring to a scenario where the monetary authority takes over market lending on behalf of the intermediaries. In the model, this corresponds to the market for differentiated capital goods. Importantly, this policy alters the distribution of marginal costs in the banking sector. We assume that the cost of funds of the central bank is lower than of any bank in the ergodic distribution. Effectively, the central bank is offering claims to firms at a subsidized price. Firms that borrow from banks with ex-ante high marginal cost are the biggest beneficiaries of this policy. The central bank is not balance sheet constrained and the subsidized lending market doesn't face any adverse selection frictions. We also assume that there is no crowding out of existing lending of other market participants.

We consider *targeted* direct lending and run decile-specific policy counterfactuals, similarly to the way we approached direct equity injections. We thus assume that the central bank takes over loan books of individual banks. Our objective, as before, is to compute conditional elasticities of aggregate output with respect to these targeted policy interventions. Formally, we compute the following object in the model:

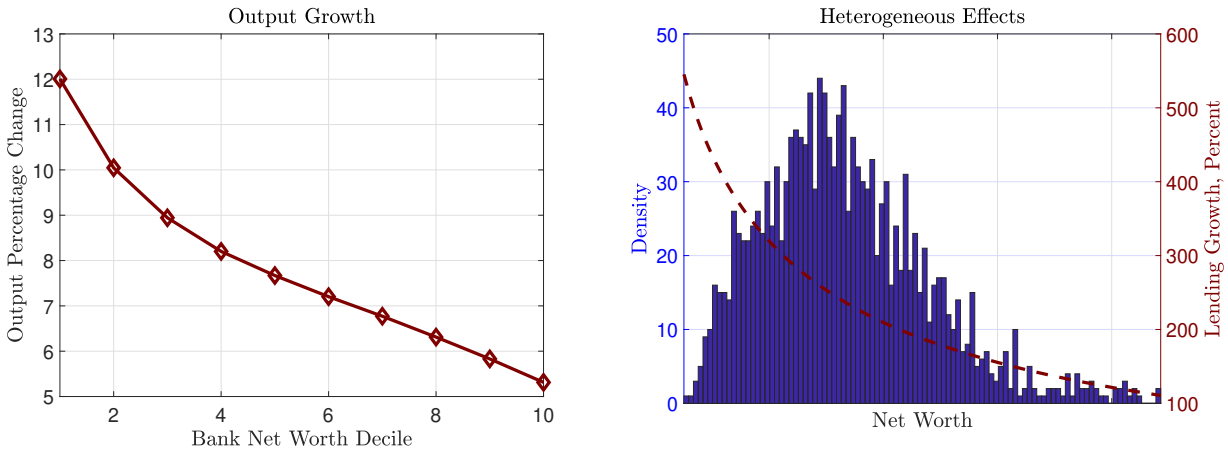
$$\frac{\partial \mathrm{Y}}{\partial \hat{\mathrm{P}}} = \frac{\partial \mathrm{Y}}{\partial \hat{\mathrm{K}}} \times \int_{\mathbf{B}} \frac{\partial \mathrm{k(j)}}{\partial \hat{\mathrm{p}}(\mathrm{j})} \mu(\mathrm{dn}, \mathrm{d}\xi) \tag{45}$$

where the left-hand side is the response of aggregate output to the direct lending facility introduction, captured by $\hat{\mathrm{P}}$. The distribution of $\hat{\mathrm{p}}(\mathrm{j})$ reflects new relative prices after the central bank's involvement with the credit portfolio of a particular institution (j). Specifically, for each decile of the distribution, we assume that the new relative price $\hat{\mathrm{p}}(\mathrm{j})$ is consistent with the lowest marginal cost available in the state space.[16] On the right-hand side we show the decomposition into the marginal product of capital and the distribution of bank-level changes in lending with respect to the new, central bank-generated relative prices. The central bank is thus essentially indirectly adjusting the market's credit supply elasticity.

Results of this policy experiment are plotted on Figure 23. On the right panel we see the stationary distribution of net worth from the baseline economy overlaid with the bank-level growth in lending in response to the policy intervention. Notice how the schedule is *decreasing* in bank net worth, because the marginal cost is lower for larger banks. The direct lending policy scheme generates a relative price advantage only for the low-net worth institutions whose cost of funds is high ex-ante. This is in direct contrast to the effects of direct equity injections. On the left panel of the Figure, we compute the elasticity of aggregate output with respect to decile-specific direct lending interventions, which is intuitively downward-sloping. Direct lending interventions targeted at smaller institutions are more effective at stimulating lending and demand.

[16] Technically, the state space is larger than the ergodic distribution. For this reason, even the largest banks in the distribution increase lending in response to this policy. Alternatively, we can normalize the $\hat{\mathrm{p}}(\mathrm{j})$ to be consistent with the lowest marginal cost in the ergodic distribution. Results and conclusions would not change.

Figure 23: **Macroeconomic Effects of Targeted Direct Lending Facilities**



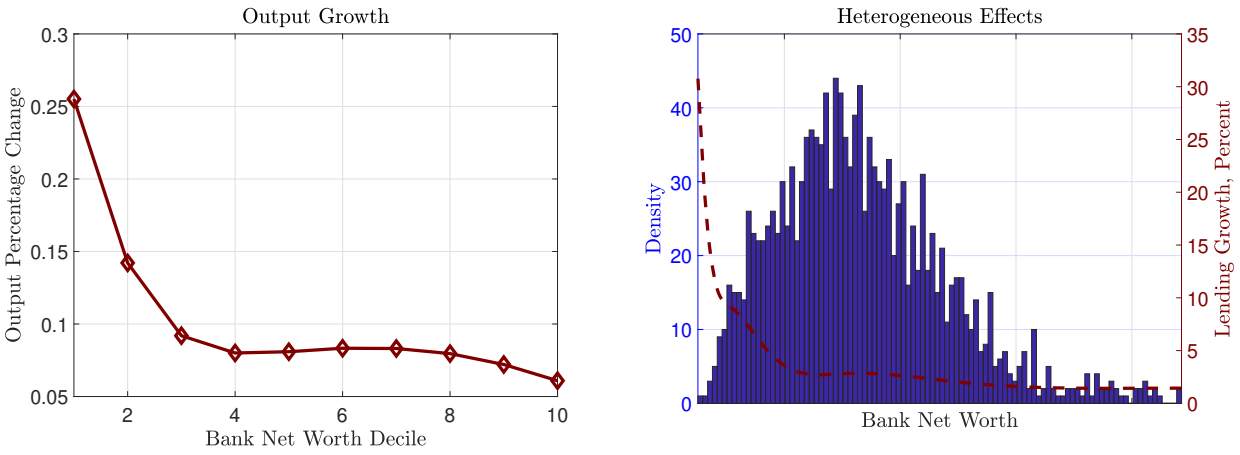
Notes: Responses of aggregate output to targeted, decile-specific direct lending facilities.

C.2 Bank-Level Debt Guarantees

We now consider *targeted* deposit insurance transfers that compensate the household for whatever losses any given bank in the distribution could deliver ex-post. In other words, if n(j) turns negative and the realized deposit recovery rate is less than unity, the government steps in on behalf of the bank and repays the promised volume of deposits in full. We assume that there are no coverage limits and that transfers are funded by lump-sum taxes on the household ex-post. This exercise supplements the deposit insurance scheme from Section 4.2 which was an aggregate policy.

Recall that introduction of deposit guarantees shuts down one source of marginal cost heterogeneity that stems from deposit rates. That is, there is an explicit guarantee that no bank in the distribution can ever default ex-post even though default risk is internalized ex-ante. This immediately implies that there is no deposit rate heterogeneity as all banks can borrow at the risk-free rate - there is no deposit spread. Recall that as long as markets are incomplete, banks still face idiosyncratic returns ξ(j), which feeds into balance sheet heterogeneity ex-post. Just as before, we segment the distribution of bank net worth into 10 deciles $\iota = 1 \ldots 10$. For each ι, we shut down the idiosyncratic default risk channel, which lowers the cost of funds of all banks in that decile to the risk-free rate. We then compute the MPL(j) and the macro elasticities subject to the new conditional distributions ten separate times, one per each affected decile.

Figure 24: **Macroeconomic Effects of Targeted Deposit Guarantees**



Notes: Responses of aggregate output to targeted, decile-specific deposit guarantees.

Figure 24 presents the result. The right panel shows how banks of different levels of net worth respond to the introduction of the debt guarantee scheme. The left panel plots the macro elasticity schedule. We see from both panels that deposit guarantees increase bank risk-taking and credit supply, which leads to output growth. Quantitatively, the level effect is rather small, if compared to other credit policies that we considered, but generally depends on the calibration of σ_ξ. In terms of heterogeneous effects, deposit guarantees increase lending of small banks by more. The distribution of deposit rates is concentrated in the left tail of the net worth density. Elimination of this channel therefore mechanically affects only the smallest intermediaries, which explains the slope of the curve.

D MIT Shocks to Aggregate Productivity

In this section we study the transmission mechanism of exogenous, unanticipated aggregate shocks to Total Factor Productivity (A_t). After a sudden one standard-deviation decline, A_t reverts back to the steady state with an autoregressive factor of 0.6. We assume that any policy interventions are fully unanticipated and occur only during crisis episodes and never in the steady state or when productivity is high.

We are interested in tracking the responses of all aggregate quantities and prices but focus on aggregate demand K_t for compactness. Let us write K_t as an explicit function of the exogenous transitory shock, equilibrium prices, and policy interventions $\{\Omega_t\}_{t\geq 0}$, with $\{\Omega_t\} = \{R_t^k, \bar{R}_t, P_t, \tau_t\}$ and where τ_t summarizes any policy actions of the government:

$$K_t\Big(\{\Omega_t\}_{t\geq 0}\Big) = \int k_t\Big(n, \xi; A_t, \{\Omega_t\}_{t\geq 0}\Big)\mu_t(dn, d\xi) \tag{46}$$

where $k_t\Big(n, \xi; A_t, \{\Omega_t\}_{t\geq 0}\Big)$ is the bank-level policy function for bank credit (assets). Recall that $\mu(n, \xi)$ is the joint distribution of bank net worth and idiosyncratic rate of return risk.

We can decompose the total response of credit supply at $t = 0$ by differentiating Equation 46:[17]

$$dK_0 = \Big[\underbrace{\int_0^\infty \frac{\partial K_0}{\partial A_t} dA_t}_{\text{Direct Effect}} + \int_0^\infty \underbrace{\Big(\frac{\partial K_t}{\partial \bar{R}_t} d\bar{R}_t + \frac{\partial K_t}{\partial R_t^k} dR_t^k + \frac{\partial K_t}{\partial P_t} dP_t + \frac{\partial K_t}{\partial \tau_t} d\tau_t\Big)}_{\text{Indirect Effect}}\Big] dt \tag{47}$$

The first term in Equation 47 summarizes direct effects of the shock to productivity on credit supply, while holding all aggregate prices and policies constant. All banks in the distribution respond to A_t directly because aggregate productivity impacts the path of aggregate returns on investment, which enters explicitly the law of motion of bank net worth through $R_t^T(j)$. The direct effect can be further decomposed into the cross section of bank-level marginal propensities to lend:

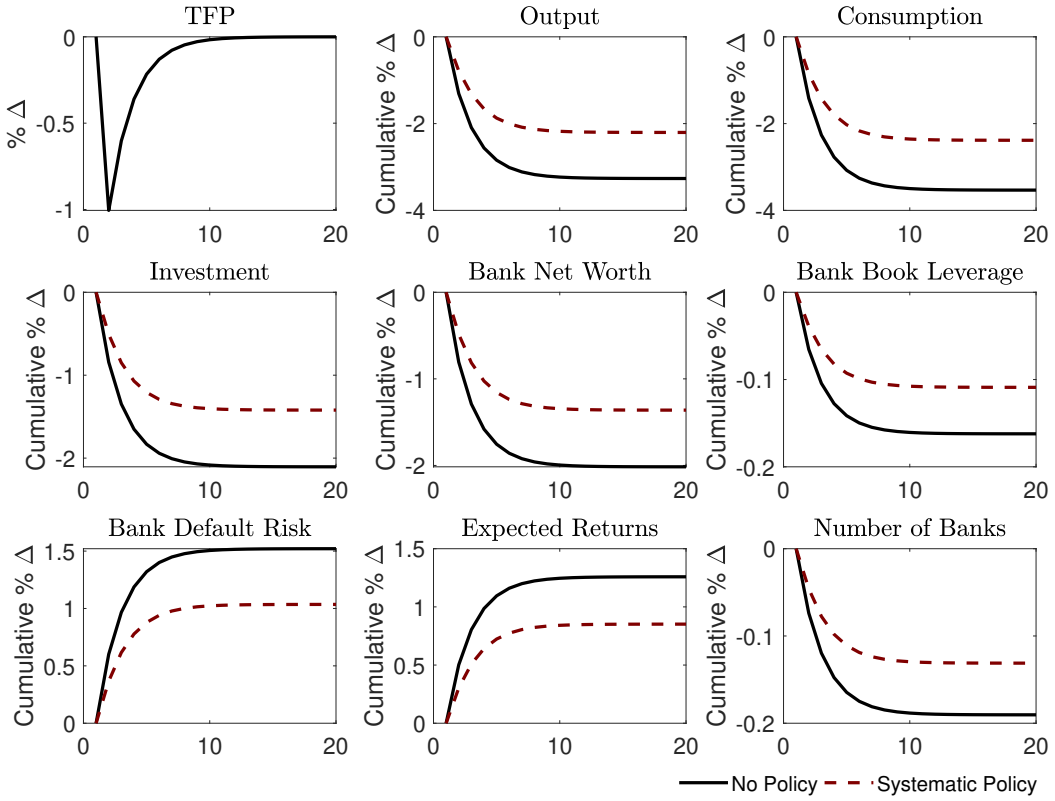
$$\int_0^\infty \frac{\partial K_0}{\partial A_t} dA_t dt = \int_0^\infty \Big[\int \frac{\partial k_0(n, \xi; A_t, \{\bar{\Omega}_t\}_{t\geq 0})}{\partial A_t} \bar{\mu}(dn, d\xi)\Big] dA_t dt \tag{48}$$

With $\bar{\Omega}$ and $\bar{\mu}$ fixed at the steady-state values. That is, the total direct effect comprises the aggregated partial-equilibrium response of credit supply to the exogenous disturbance alone without updating aggregate general equilibrium variables and the banking distribution.

The indirect effect from Equation 47 includes four distinct channels of transmission. First, aggregate productivity impacts demand for investment. Because firms require external financing in order to produce, this immediately translates into the demand for bank lending activities. Banks, because of credit market power, respond to increased demand by adjusting their private, bank-level markups and prices. In addition, prices adjust also because the distribution of bank net worth shifts and, as we concluded in previous sections, relative prices and marginal costs vary with net worth. In the aggregate, this moves P_t, which further feeds into bank-level choices of credit supply. Recall that banks do not internalize this GE channel, which is an aggregate credit supply externality.

[17] Our decomposition is similar to the one applied in Kaplan et al. (2018) and Auclert (2019) who study distributional implications in the response of aggregate consumption to monetary policy shocks.

Figure 25: **Crisis Experiment: Systematic Equity Injections**

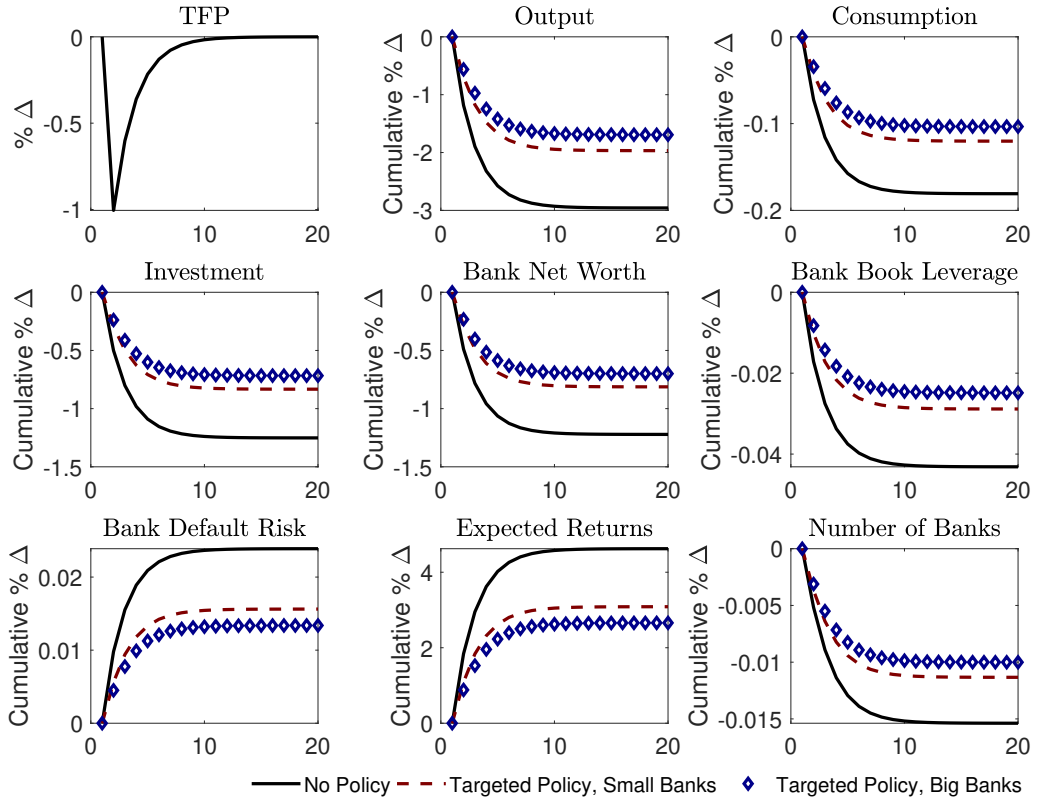


Notes: Responses to a one standard-deviation negative shock to A_t, with and without systematic equity injections. Baseline economy.

Second, banks react to movements in investment demand by requesting more or less short-term debt from the households. In the deposit market, this affects the distribution of deposit interest rates, which drives the aggregate rate $\bar{R}_t$. Third, every second-level general equilibrium channel feeds into the aggregate stock of capital which, together with the aggregate price, determines the new level of systematic returns R_t^k.

Finally, banks will react to policy interventions from the fiscal and monetary authorities, if there is any. In previous sections, we discussed systematic and targeted (bank-level) equity injections as well as lending and liquidity facilities, and deposit guarantees. All these policies are summarized in the term τ_t, which is understood to be capturing any aggregate or bank-level policy responses. Credit policies of any kind will perturb allocations in the banking sector one way or another. Equity injections induce direct credit supply responses because those explicitly augment one of the idiosyncratic states of the banking problem - $n_t(j)$. Lending facilities shift the distribution of relative prices, thus affecting lending decisions both directly and indirectly through the aggregate price of capital P_t. Liquidity facilities impact the probability of the leverage constraint binding in the future, which weighs in on the banks' decision to take on more or less balance sheet risk. Finally, deposit guarantees shift the cost of capital schedule and have both direct and indirect (through interest rates on deposits) effects on bank lending.

Figure 26: **Crisis Experiment: Targeted Equity Injections**

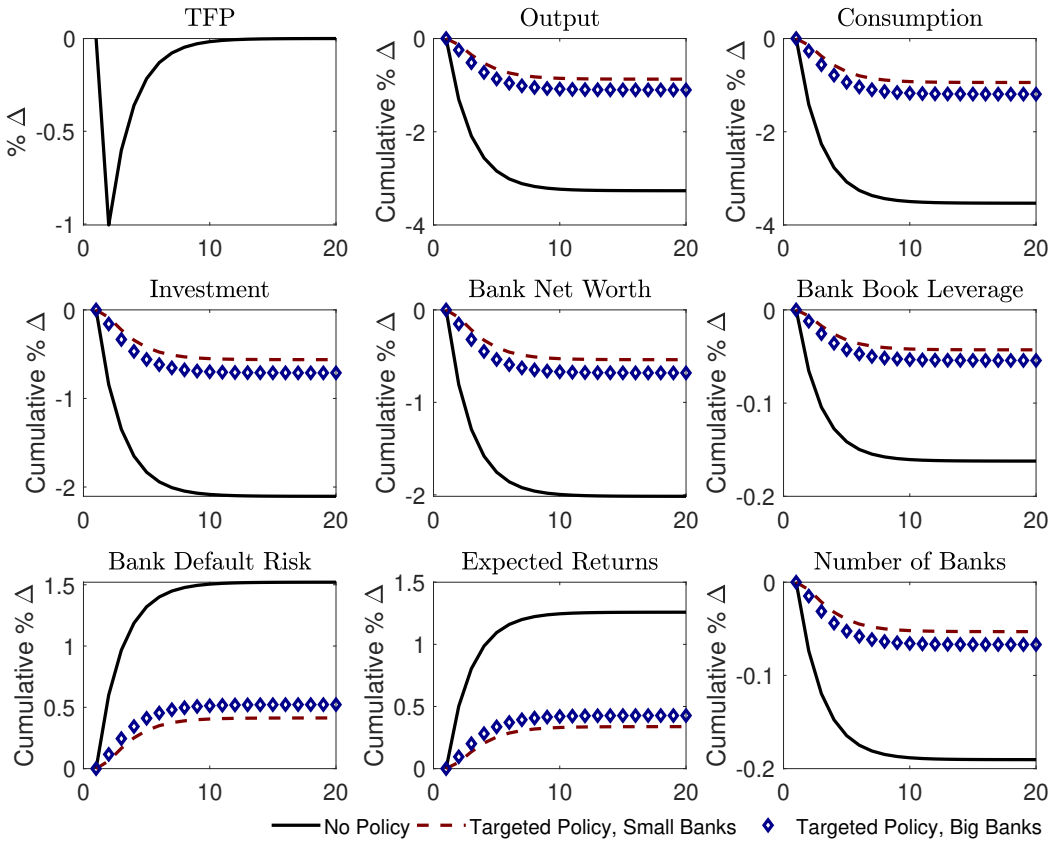


Notes: Responses to a one standard-deviation negative shock to A_t, with and without targeted equity injections. Baseline economy.

We begin the presentation of numerical results with our baseline economy that goes through an aggregate economic crisis but gets a response from fiscal/monetary institutions in the form of systematic equity injections into the banking industry. We assume that the government increases net worth of every bank in the distribution by 10%. Figure 25 portrays the results. For all variables except the transitory A_t shock, we present cumulative impulse response functions in order to best illustrate the efficacy of this policy. We observe that the economy is going through a contraction of aggregate consumption, output, and investment of the magnitudes that are similar to the 2007-2009 Great Recession. Bank net worth and book leverage fall. Bank balance sheets become more risky as the aggregate (average over the entire distribution) probability of insolvency risk increases. As the average bank in the distribution is smaller in terms of net worth, the leverage constraint binds or is more likely to bind for a larger fraction of the intermediaries. This translates into the rise of equilibrium excess returns. Finally, because bank franchises decline in value, fewer banks decide to enter and the number of active intermediaries falls.

Figure 25 also plots impulses and responses under the government's systematic equity injections strategy. Recall that this policy entails increasing net worth of all banks in the distribution by 10% relative to the pre-crisis steady-state levels. We see that policy has a significant supporting effect on all macroeconomic aggregates. A boost of net worth induces a smaller reduction in credit supply

Figure 27: **Crisis Experiment: Liquidity Facility**



Notes: Responses to a one standard-deviation negative shock to A_t, with and without targeted liquidity facilities.

and the probability of insolvency. Leverage constraints are less likely to bind now or in the future, thus the dampening impact on excess returns. The number of active intermediaries falls by less because bank franchise values are buttressed by government funds, which in turn induces more entry than in the no-policy scenario. Overall, these results are similar to the discussion in Gertler and Kiyotaki (2010) who conduct a similar experiment but in the environment with complete markets, perfect competition, and exogenous entry.

In Figure 26 we begin to consider targeted, bank-level policy interventions. First, we look at direct equity injections into only small or only large banks, and compare the response functions. We define "small" and "large" banks as those intermediaries whose net worth is in the bottom and top deciles of the stead-state, ergodic distribution of bank net worth, respectively. For each case, we consider the same 10% government-sponsored increase in equity. We see that equity that is injected into big banks has a bigger bang for the buck than the equivalent investment into small banks. This is due to the positive slope of the MPL schedule and a bigger macro elasticity. Large banks have a greater propensity to lend than small banks because of lower marginal costs and increasing returns to scale (with respect to total cost of capital, including non-interest and interest expenses). Quantitatively, the difference between small-bank and big-bank policies is not

negligible.

Figure 27 plots the same numerical experiment but now with targeted liquidity facilities. Recall that these policies reduce the fraction of divertible assets λ by 10% for a certain decile of the bank net worth distribution. We see that discount-window-based lending considerably dampens recessions, particularly if applied to small banks. This is due to the leverage constraint binding much more often for banks with low levels of net worth, particularly in recessions when net worth is low. Thus, any policy that reduces λ induces a greater credit supply response if directed to the agents that are affected by the moral hazard friction by more. Because now the risk of a tightening leverage constraint relatively dissipates, excess returns increase by less, which leads to lower risks of default, more lending, and a relatively stronger macroeconomic responses (the economy is still contracting but the cumulative magnitude is lower).

E Solution Algorithm

In this section we lay out the numerical algorithm that is used to solve different variants of the model. Section E.1 describes how to solve the baseline unregulated market economy. Section E.3 solves the constrained efficient allocations of the social planner. Finally, E.3 solves the regulated economy which decentralizes the social planner's solution.

E.1 Unregulated Market Equilibrium

There are four basic computational challenges when solving the model in Section 2. First, we must solve the dynamic optimization problems of the financial intermediary and of the household. Because the banking sector is not scale invariant, individual bank characteristics matter for aggregation. The individual state vector is $\{n(j), \xi(j)\}$. We use value function iteration to solve the banking problem. For the household's problem we use time iteration on the household's Euler equation for deposit holdings.

Second, banks face an occasionally binding constraint on leverage. We deal with this issue the following way. On each point of the idiosyncratic state space we solve for optimal balance sheet choices while assuming the constraint binds. Then, we back out the Lagrange multiplier on the constraint. If the constraint is, in fact slack, we solve the problem again using numerical global maximization routine. If it is binding, then we continue the program.

Third, the market for deposit holdings must clear between bankers and the household. Because there is no deposit insurance, households price the full distribution of bank default risk into the menu of deposit rates. We iterate on the equilibrium deposit rates in the outer loop of the program. In each iteration, the newly constructed household's stochastic discount factor Λ - an endogenous aggregate state - is adopted by the banks.

Finally, there are 2 key aggregate (static) variables that are needed to pin down R^k: K and P. We use a variant of the stochastic simulation approach as in Maliar et al. (2010) to pin them down. Specifically, after solving for both the household's and banker's problems, we run a long simulation with N=1 intermediary and T=20,000 periods while using the newly computed policy functions of the incumbent banker. The distribution grants us new measures of aggregate demand of the incumbent. We combine these with demand from new entrants (whenever entry is endogenous), which gives us new candidates for aggregate capital and prices. We require convergence tolerances of 10^{-6} for general equilibrium deposit rates, and 10^{-5} for the bankers' and household value functions, and 10^{-3} for aggregate capital and prices.[18] Below we list the state variables of the model and sketch the solution algorithm.

Exogenous idiosyncratic shocks: $\{\xi(j)\}$. Exogenous idiosyncratic states: $\{n(j)\}$

Endogenous idiosyncratic states: $\{\upsilon(j), \bar{R}(j)\}$. Endogenous aggregate states: $\{K, P, \Lambda\}$

Algorithm - Stationary Industry Equilibrium

1. Guess some initial values for aggregate endogenous states $\{K, P, \Lambda\}$. Compute R^k. Guess some initial values for idiosyncratic endogenous states $\{\upsilon(j), \bar{R}(j)\}$

[18] Importantly, there is no aggregate risk in the model. We therefore do not need to track a dynamic distribution of bank net worth in the present paper.

2. Solve the financial intermediation problem

   (a) Use value function iteration. On each grid point, assume the leverage constraint binds.

   (b) Construct the Lagrange multiplier. If constraint indeed binds, proceed.

   (c) If constraint is slack, solve the problem again using a numerical minimization routine.

3. Simulate the problem of the incumbent. Run a simulation of N=1 bankers and T=20,000 periods.

4. Solve the new entry problem, if entry is endogenous. Determine the mass of entrants and their aggregate demand for capital in each period of the simulation.

5. Compute economywide new guesses for aggregate K' and P'. Construct a new $R^{k'}$. Check if K' is sufficiently close to K. If not, return to Step 2. If yes, continue to the program.

6. Calculate the probability of bank default on each grid point using the newly computed policy functions and distributional aggregates. This gives new $\{v'(j), \bar{R}'(j)\}$.

7. Solve the household's problem. Get new Λ'.

8. Compare $\{\bar{R}(j)\}$ with $\{\bar{R}'(j)\}$. If maximal error cross all grid points is within the tolerance level, general equilibrium is solved. If not, update $\{\bar{R}(j)\}$. Return to Step 2 and continue the iteration.

E.2 Constrained Efficient Equilibrium

In order to solve for constrained efficient (socially optimal) allocations, we must make one adjustment to the algorithm. The only difference between the decentralized solution and the social planner is that the latter internalizes the impact of private choices on aggregate returns. We operationalize this using projection methods. Specifically, we assume that both K and P are polynomials in n(j), $\xi(j)$, and the choice of k(j). That is:

$$\begin{aligned} K &= \alpha_0^k + \alpha_1^k n(j) + \alpha_2^k \xi(j) + \alpha_3^k k\Big(n(j), \xi(j)\Big) \\ P &= \alpha_0^p + \alpha_1^p n(j) + \alpha_2^p \xi(j) + \alpha_3^p k\Big(n(j), \xi(j)\Big) \end{aligned}$$

Once the optimal k(j) is found, that gives us p(j) through the credit demand function and d(j) from the balance sheet constraint. The objective of the projection is then to find the optimal vector of coefficients $\{\alpha^k, \alpha^p\}$. We now describe the steps of the algorithm below.

Algorithm - Constrained Efficient Equilibrium

1. Guess some initial values for $\{\alpha^k, \alpha^p\}$.

2. Guess some initial values for aggregate endogenous states $\{K, P, \Lambda\}$. Compute R^k. Guess some initial values for idiosyncratic endogenous states $\{v(j), \bar{R}(j)\}$. Decentralized equilibrium solution works as a good first guess

3. Solve the financial intermediation problem under the social planner

   (a) Given $\{\alpha^{k}, \alpha^{p}\}$, treat R^{k} as endogenous to the states and to the candidate choices of k(j). Use a numerical minimization routine to solve for the optimal k(j) on each grid point.

   (b) On each grid point, first assume the leverage constraint binds.

   (c) Construct the Lagrange multiplier. If constraint indeed binds, proceed.

   (d) If constraint is slack, solve the problem again using a numerical minimization routine. Keep treating R^{k} as endogenous to states and choices.

4. Simulate the problem of the incumbent. Run a simulation of N=1 bankers and T=20,000 periods. Run a linear regression of capital holdings k(j) on a constant, lagged net worth $n_{t-1}(j)$, lagged $\xi_{t-1}(j)$, and lagged capital holding $k_{t-1}(j)$. Do the same for p(j). Compute new guesses for $\{\alpha^{k'}, \alpha^{p'}\}$.

5. Solve the new entry problem, if entry is endogenous. Determine the mass of entrants and their aggregate demand for capital.

6. Compute economywide new guesses for K′ and P′. Construct a new $R^{k'}$. If K' and P' are sufficiently close to K and P, respectively, then continue. If not, return to Step 2.

7. Calculate the probability of bank default on each grid point using the newly computed policy functions and distributional aggregates. This gives new $\{v'(j), \bar{R}'(j)\}$

8. Solve the household's problem. Get new Λ'.

9. Compare $\{\alpha^{k}, \alpha^{p}\}$ with $\{\alpha^{k'}, \alpha^{p'}\}$. And compare $\{\bar{R}(j)\}$ with $\{\bar{R}'(j)\}$. If the maximal errors across all grid points are within the tolerance level, the constrained efficient equilibrium is solved. If not, update $\{\alpha^{k}, \alpha^{p}\}$ and $\{\bar{R}(j)\}$. Return to Step 3 and continue the iteration.

E.3 Decentralization

We decentralize constrained efficient equilibria with size-dependent taxes on bank gross returns. In our procedure, we want the decentralized solution to converge to the social planner's allocations both in terms of aggregates and in the banking cross-section. We impose on the decentralized solution optimal endogenous aggregates states from the constrained efficient equilibrium above. Then, we iterate on a tax schedule until all policy functions yield idiosyncratic endogenous states (the deposit rate schedule) that are exactly consistent with the social planner's allocations. Specifically, we start with a guessed $\tau(j)$ for each grid point. We solve the problem of the financial intermediaries subject to the tax schedule and compute a new guess for $\bar{R}(j)$, and so on until convergence. We do not update the household's solution or run simulations in the intermediate step, because the aggregate endogenous states are fixed. The exact algorithm is described below

Algorithm - Regulated Market Equilibrium

1. Start with the solution to the decentralized equilibrium. Impose new aggregate endogenous states that are now permanently equal to the social planner's values: $\{K_{sp}, P_{sp}, \Lambda_{sp}\}$. Compute R^k once and do not change during the iteration.

2. Guess some initial values for the gross return tax $\tau(j)$

3. Solve the financial intermediation problem

   (a) Treat return taxes $\tau(j)$ as given

   (b) Use value function iteration. On each grid point, assume the leverage constraint binds.

   (c) Construct the Lagrange multiplier. If constraint indeed binds, proceed. If constraint is slack, solve the problem again using a numerical minimization routine

4. Calculate the new probability of bank default on each grid point using the newly computed policy functions. This gives new $\{\upsilon'(j), \bar{R}'(j)\}$.

5. Compare the policy function for net worth n' obtained in this iteration with the social planner's solution n'_{sp}. Use a bisection method for constructing a new candidate $\tau'(j)$. if $n'(j) > n'_{sp}(j)$ for any j, increase the tax rate on that grid point. Alternatively, the regulated solution is too small and we decrease the tax rate on that grid point.

6. Compare the new $\tau'(j)$ with the old $\tau(j)$. If the maximal squared error across all grid points is within the tolerance level, complete the program. Alternatively, update $\tau(j)$ and revert back to Step 3.

UniCredit Foundation
Piazza Gae Aulenti, 3
UniCredit Tower A
20154 Milan
Italy

Giannantonio De Roni – *Secretary General*
e-mail: giannantonio.deroni@unicredit.eu

Annalisa Aleati - *Scientific Director*
e-mail: annalisa.aleati@unicredit.eu

Info at:
www.unicreditfoundation.org


UniCredit Foundation