Bank of Japan Working Paper Series

Interest Rate Pass-through by U.S. Banks: Macro Implications of Bank Competition

Nobuhiro Abe*
nobuhiro.abe@boj.or.jp

Yuto Ishikuro**
yuuto.ishikuro@boj.or.jp

Koki Nakayama***
kouki.nakayama@boj.or.jp

Yutaro Takano**
yuutarou.takano@boj.or.jp

No.25-E-9
September 2025

Bank of Japan
2-1-1 Nihonbashi-Hongokucho, Chuo-ku, Tokyo 103-0021, Japan

* Financial System and Bank Examination Department
(currently Monetary Affairs Department)
** Financial System and Bank Examination Department
(currently Personnel and Corporate Affairs Department)
*** Financial System and Bank Examination Department
(currently Research and Statistics Department)

Interest Rate Pass-through by U.S. Banks: Macro Implications of Bank Competition [*]

Nobuhiro Abe,[†] Yuto Ishikuro,[‡] Koki Nakayama,[§] and Yutaro Takano[¶]

September 2025

Abstract

Do heterogeneity and competition among banks matter for the macroeconomy? To address this question, we develop a Heterogeneous Bank New Keynesian (HBANK) model that incorporates oligopolistic competition among banks in both loan and deposit markets into an otherwise canonical New Keynesian model. We calibrate model parameters for the cost structure and demand for loans and deposits using data of the 170 largest banks in the U.S. Differences in the parameter values reflect differences among banks in the size of duration risk they take, markups of loan rates, and markdowns of deposit rates. Based on simulation exercises, we show that aggregate lending becomes more responsive to monetary and productivity shocks in our HBANK model than in a Representative Bank New Keynesian model (RBANK), primarily because of heterogeneity in duration risk and the responsiveness of loan markups among banks.

Keywords: Banking, Business Cycles
JEL classification: E32, E43, E44, E52, G21

[*]We would like to thank Kosuke Aoki, Wataru Hirata, Takeo Hoshi, Toshitaka Maruyama, Fumitaka Nakamura, Mitsuhiro Osada, Nao Sudo, and the staff of the Bank of Japan for their valuable comments and discussions. The views expressed in this paper are those of the authors and do not necessarily reflect the official view of the Bank of Japan or of the Financial System and Bank Examination Department. Any remaining errors are those of the authors.

[†]Email: nobuhiro.abe@boj.or.jp
[‡]Email: yuuto.ishikuro@boj.or.jp
[§]Email: kouki.nakayama@boj.or.jp
[¶]Email: yuutarou.takano@boj.or.jp

1 Introduction

The presence of a banking sector is of great importance for the transmission of monetary policy and other structural shocks. Indeed, the macro-banking literature has developed various models that incorporate financial frictions and has long highlighted their importance in amplifying and propagating macroeconomic changes (e.g. Bernanke and Blinder (1988), Bernanke and Gertler (1995), and Gertler and Kiyotaki (2010)). However, macroeconomic analysis of bank heterogeneity and competition is still limited and how these factors matter for business cycle fluctuations is still an open question. By contrast, the literature has clearly noted the presence of heterogeneity in the banking system both empirically and theoretically (e.g. Kashyap and Stein (2000) and Drechsler et al. (2017)). Figure 1 shows the development of two measures of concentration in the loan and deposit markets in the U.S. over the last three decades. The market share of the top 100 banks in both markets has increased from 50% to 80%.

In this paper, we develop a model which we refer to as the "Heterogenous Bank New Keynesian (HBANK) model" by incorporating oligopolistic competition in both deposit and loan markets into an otherwise canonical macro-finance model, and study the effects of heterogeneity and competition among banks on the propagation of macroeconomic shocks to the aggregate economy. Our HBANK model is built on two prior studies - Abadi, Brunnermeier, and Koby (2023), which considers imperfect competition among banks in deposit and loan markets, and Atkeson and Burstein (2008), which considers oligopolistic competition in which firms' markups change with their market shares. We apply the framework of Atkeson and Burstein (2008) to the deposit and loan markets of the model considered in Abadi, Brunnermeier, and Koby (2023). In addition to varying in their preferences for duration risk, banks are by their nature different in terms of their efficiency of producing services. This difference results in endogenous differences in banks' markups and markdowns, as well as asset sizes and asset portfolios. We calibrate the parameters associated with banks' loan and deposit

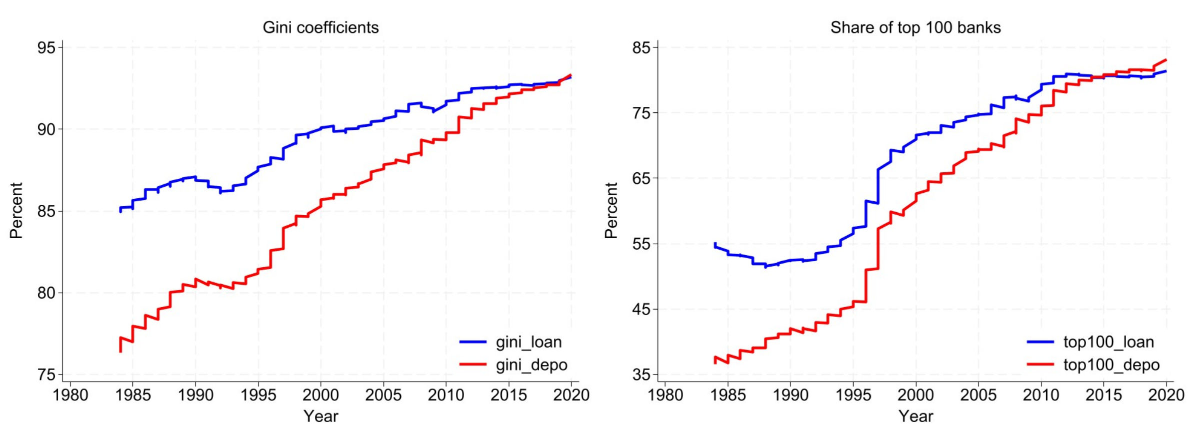

Figure 1: Time series of Gini coefficiens and market share of top 100 US banks in loan and deposit markets

production functions and those associated with the demand for banks' loans and deposits using data of individual U.S. banks.

Next, we compare the behavior of the HBANK model with the Representative Bank New Keynesian (RBANK) model, in which bank heterogeneity is absent, to analyze the role of banks' heterogeneity in macroeconomic responses to exogenous shocks, such as monetary and productivity shocks. In addition, we perform Bayesian estimation of the model using time series data and present variance decomposition of key macroeconomic varaibles, such as GDP and investment, to analyze the contributions of various shocks, including shocks to the aggregate banking sector, to variations in these variables.

In the first part of the paper, we show that our HBANK model reasonably replicates the distribution of deposit and loan rates in terms of their levels and degree of pass-through, as well as the market shares of deposits and loans of each bank. Over the period examined, large banks with large market shares of deposits and loans set low deposit rates and lending rates compared with smaller banks. Moreover, when the Federal Reserve Board raises the federal funds rate, the increase in the lending rate tends to be large, while the increase in the deposit rate tends to be relatively modest for large banks compared to smaller banks (e.g. English et al. (2018), Drechsler et al. (2021), and Gomez et al. (2021)). Our model replicates these micro observations in the sense that at the steady state larger banks have lower deposit and loan rates than smaller banks and, around the steady state, they exhibit larger responses of loan rates and smaller responses of deposit rates to short-term interest rate shocks than smaller banks.

Two points are noteworthy regarding how our model replicates the data. First, we introduce what we refer to as a quality adjustment shifter into Atkeson and Burstein (2008)'s nested CES function. The quality adjustment shifter represents the quality of deposit and/or loan services provided by a bank and plays an important role in shifting relative demand for deposits and/or loans across banks. In the data, the deposit rates of large banks are lower than those of small banks, while large banks retain a large share of the deposit market. This suggests that there are quality differences in deposit services among banks that lead depositors to put their deposits into large banks even though they are offered lower deposit rates. Indeed, by calibrating the model with the quality adjustment shifter using U.S. Call Report data, we show that large banks offer both high quality loans and deposits. We also show that due to differences in the quality of loan and deposit services, large banks charge a higher markup in the loan market and a lower markdown in the deposit market (larger deposit spread). These observations are in line with key findings in the literature (e.g. Corbae and D'Erasmo (2021) and Jamilov and Monacelli (2023)).

Second, we incorporate banks' duration risk, following the discussion in Drechsler et al. (2021). Indeed, Drechsler et al. (2021) finds that income β^{Inc} is larger for banks with more assets and shorter asset duration in the U.S. The differences in risk-taking behavior among banks play the critical role in determining the differences in the degree of pass-through of short-term interest rates to lending rates, β^{Inc}, of banks and interest income of banks. In particular, in our model, large banks with short duration benefit from a larger increase in interest income when the market interest rate increases (high β^{Inc}).[1]

Using the calibrated model, we show that the impacts of structural shocks, such as monetary policy shocks and productivity shocks, on GDP or corporate investment are larger in the HBANK model than in the RBANK model. In the HBANK model, heterogeneity in interest rate setting and asset allocation behavior among banks (a higher loan markup or a shorter loan duration among large banks, for example) helps amplify the propagation of exogenous shocks to the rest of the economy. Moreover, heterogeneous loan pass-through of banks' loan rates leads to a shift in loans outstanding across lending banks and this contributes to a sizable impact on the aggregate loan interest rate and the aggregate demand for loans. More specifically, in the case of an interest rate hike shock, the demand for loans shifts from large banks, which have relatively large increases in lending rates, to small and medium-sized banks, which have limited increases in lending rates. A shift in borrowing demand to small and medium-sized banks, which have higher lending rates in the steady state, causes aggregate lending rates to rise more than in the RBANK model. Thus, in the HBANK model, a higher lending rate leads to a larger contraction of corporate investment, which leads to a larger decline in GDP. This implies that a minor change in the policy rate or productivity level may cause a larger change in the aggregate economy than what conventional wisdom suggests. This volume-shift channel plays an analogous role in the deposit market.

Finally, in order to analyze the implications of bank heterogeneity from a macro perspective quantitatively, we perform Bayesian estimation for both the HBANK and the RBANK models using time series of macroeconomic data for the U.S. Specifically, we estimate nine macro shocks, including traditional shocks such as monetary, productivity, and markup shocks, rather than idiosyncratic shocks to individual banks. Adding to the conventional shocks, we also estimate *macro* duration-risk shocks in order to examine how heterogeneity among banks affects the propagation mechanism of duration-risk shocks.[2] Our results are twofold. First, we find that duration-risk shocks do matter in explaining business cycle fluctuations of GDP and corporate investment. Second, we find that productivity shocks explain only a small fraction of the variation in GDP with the RBANK model. The contribution of productivity shocks is three times as large in the HBANK model. This finding underscores the importance of banks' heterogeneity when quantifying the size of structural shocks hitting the financial system and the economy and evaluating the relative significance of such structural shocks, as heterogeneity could amplify the propagation mechanism of a particular size and

[1] Admittedly, duration risk is considered essential for understanding the effect of unconventional monetary policy as well. In gereral, unconventional monetary policy, such as quantitative easing, acts to suppress long-term yields. In the low interest rate environment, banks may search for yield and take duration risk (Buch et al. (2014)). It is important to examine this mechanism when considering the bank lending channel, as shown in our analysis.

[2] We use this duration-risk shock to evaluate the effect of banks' preferences over asset duration. The time-variation of duration risk at the aggregate level of the banking sector could impact the macroeconomy through changing the size and/or the composition of bank portfolios Drechsler et al. (2021).

type of shock in an important manner.

Literature review Our research is part of the long-standing literature on the bank lending channel that explores the impact of monetary policy on lending through bank balance sheets (e.g. Bernanke and Blinder (1988), Bernanke and Gertler (1995)). Our model is a "macro-banking" model that incorporates the financial intermediation sector into a macroeconomic framework. Our paper therefore relates to papers quantifying the impact of financial frictions on the dynamics of the macroeconomy (e.g. Gertler and Kiyotaki (2010) and Gertler and Karadi (2011), He and Krishnamurthy (2013), Brunnermeier and Sannikov (2014), and Drechsler et al. (2018)).

Among the various types of frictions, this paper focuses on the importance of banks' market power in deposit and loan markets. Banks' market power has long been studied, beginning with Klein (1971) and Monti et al. (1972). Our model builds on Abadi, Brunnermeier, and Koby (2023), a pioneering study which analyzed the presence of reversal interest rates theoretically. Abadi, Brunnermeier, and Koby (2023) constructs a macro-banking model, which incorporates the market power of banks in both loan and deposit markets as well as leverage and liquidity constraints on bank balance sheets.[3] The empirical study of bank market power in the credit market was pioneered by Boyd and De Nicolo (2005) and Scharfstein and Sunderam (2016). Previous empirical studies focusing on banks' market power in deposit markets include Corbae and Levine (2018), Wang (2018), and Di Tella and Kurlat (2021).

The extensive empirical applied micro literature on banking has investigated heterogeneity among banks (e.g. Kashyap and Stein (1995), Kashyap and Stein (2000), English et al. (2018), and Gomez et al. (2021)) in areas other than market power. Drechsler et al. (2017) and Drechsler et al. (2021) investigate market power by focusing on heterogeneity in the U.S. deposit market and its consequences for lending. Drechsler et al. (2017) finds that banks raising deposits in concentrated markets reduce their lending more than other banks when the federal funds rate rises (the deposit channel of monetary policy). Drechsler et al. (2021) shows that banks' net interest margins are highly stable and insensitive to interest rates, and that banks' net worth is largely insulated from monetary policy shocks. The study also shows that banks match the interest rate sensitivities of their expenses and revenues one-for-one and argues that banks with less interest-sensitive deposits hold assets with much longer duration.[4]

A literature that focuses more on the role of bank heterogeneity from a macro perspective has emerged recently. For example, Corbae and D'Erasmo (2021) builds a quantitative model that addresses imperfect competition in loan markets and finds that regulatory policy can have a substantial impact on the structure of the banking market, including changes in the allocative efficiency of re-

[3]In addition, Eggertsson et al. (2024) and Ulate (2021) present models in which the expansionary effects of monetary policy can be muted when interest rates fall below zero due to a reduction in banks' interest margins and a resulting decline in their net worth.

[4]Other studies focusing on deposit market power include Egan et al. (2017), Kurlat (2019), Whited et al. (2021), and Wang et al. (2022).

sources and banking stability. Coimbra and Rey (2024) develops a macro model with heterogeneous financial intermediaries, where heterogeneity is addressed by VaR constraints coupled with limited liability. Their model features endogenous time variation in leverage, risk-shifting, and financial stability. They argue that there should be a trade-off between economic activity and financial stability depending on the level of the interest rate. Ogawa (2025) develops a general equilibrium model where banks are heterogeneous with respect to liquidity management. In this model, smaller banks tend to accumulate capital and liquidity buffers with a precautionary motive. Using the model, the author documents that the presence of liquidity shocks were the key to understanding the large drop in the aggregate loans of banks during the Great Recession in the U.S.

Within this literature, our paper is closely related to Bellifemine, Jamilov, and Monacelli (2022) and Jamilov and Monacelli (2023). These two papers also incorporate imperfect competition in both the loan and deposit markets with the Kimball (1995) aggregator. In particular, the heterogeneous bank framework with costly bank failures of Bellifemine, Jamilov, and Monacelli (2022) generates an amplification mechanism of monetary policy shocks that is absent in the representative bank model. Our model is different from these works in two ways. First, our HBANK model incorporates oligopolistic competition among banks in the spirit of Atkeson and Burstein (2008).[5] Second, our focus is on the pass-through of interest rates in both the loan and deposit markets, and our framework focuses more on the role of duration risk that banks take rather than that of leverage constraints banks face. However, similar to Jamilov and Monacelli (2023), we find that heterogeneity among banks amplifies the response of the aggregate economy to an exogeneous shock, such a monetary or productivity shock. From this perspective, our research complements the work of Bellifemine, Jamilov, and Monacelli (2022).

Finally, our paper contributes to the literature on heterogeneous agent macro models. We apply the sequence space Jacobian to solve the model, as in Boppart et al. (2018), Auclert et al. (2021), and Auclert et al. (2020). Moreover, we perform macro-Bayesian estimation to understand the contributions of the main financial frictions, namely incomplete loan and deposit markets. To do this, we apply the Whittle approximation to the likelihood function, as in Hansen and Sargent (1981) and Plagborg-Møller (2019). This differs from the typical approach of likelihood-based estimation, which applies the Kalman filter to the model's state-space representation to compute the likelihood (e.g. Smets and Wouters (2007) and Herbst and Schorfheide (2016)) and has the advantage of allowing us to calculate the likelihood efficiently using a Fast Fourier Transform with the model's MA represenatation.

[5] Atkeson and Burstein (2008) introduces imperfect competition in the international trade market with a nested-CES demand function. Recent studies have extended this framework to analyze imperfect competition in other markets. For example, Berger et al. (2022) applies the Atkeson and Burstein (2008) framework to the U.S. labor market and estimates the within- and across-market substitution parameters in the labor market.

Structure The paper is organized as follows. In section 2, we describe the macro DSGE model that incorporates an oligopolistic competition environment into a New Keynesian model. Section 3 explains our calibration procedure for the banking system block, which uses a grid search that takes full account of the general equilibrium model and its results. In Section 4, we explore the model dynamics and study the characteristics of the HBANK model by contrasting it with the RBANK model. Section 5 performs Bayesian estimation using macro time series datasets. Section 6 concludes the paper.

2 Model

In this section, we present the model framework. The basic framework of the model relies on Abadi, Brunnermeier, and Koby (2023) (hereinafter, ABK), whose model is similar to ours in that banks have market power in both the deposit and loan markets and determine markups and markdowns based on the marginal cost of loans and the marginal return on deposits. We incorporate heterogeneity in competitiveness among banks into ABK's model of the banking system. Specifically, by introducing the framework of Atkeson and Burstein (2008)'s oligopolistic competition into the ABK model, we provide a framework for analyzing the macroeconomic impact of changes in base rates due to monetary policy on changes in deposit and lending rates under different markups and markdowns at the individual bank level.

The baseline ABK framework considers a discrete-time New Keynesian economy. The economy consists of households, intermediate goods producers, final good producers who produce a final good from intermediate goods with the CES aggregator, and capital good producers who sell the final good as investment goods, as well as the central bank, the government, and banks. The central bank conducts monetary policy by adjusing short-term and long-term interest rates. The government conducts fiscal policy by making decisions on taxes and transfers jointly. Banks set interest rates on deposits and loans in oligopolistic deposit and loan markets, where they exercise market power. Lending rates affect the economy through changes in the production costs of intermediate goods firms and deposit rates affect the economy through changes in the savings behavior of households.

We first present the New Keynesian block in section 2.1. Section 2.2 formulates the banks' problem in both deposit and loan markets and derives the optimal pricing behavior of heterogeneous banks. Section 2.3 defines the equilibrium of the economy. For a full derivation of the model results, see Appendix A for the New Keynesian block and Appendix B for the banking system block.

2.1 New Keynesian block

Households The representative household maximizes expected utility:

$$\sum_{t=0}^{\infty} \beta^t \left[e^{v_t^p} U\left(C_t, C_{t-1}, H_t\right) + e^{v_t^d} \Phi\left(D_t\right) \right],$$
$$\text{where} \quad U\left(C_t, C_{t-1}, H_t\right) = \frac{\left(C_t - hC_{t-1}\right)^{1-\sigma}}{1-\sigma} - \chi \frac{H_t^{1+\varphi}}{1+\varphi}, \quad \Phi\left(D_t\right) = -\frac{\zeta}{2}\left(D_t - D^*\right)^2. \tag{1}$$

The budget constraint of the representative household is given by:

$$C_t + B_t + D_t \leq \frac{W_t}{P_t} H_t + \frac{1 + r_{t-1}^b}{1+\pi_t} B_{t-1} + \frac{1 + r_{t-1}^d}{1+\pi_t} D_{t-1} + \mathcal{P}_t + T_t. \tag{2}$$

In the utility function, C_t represents consumption, H_t represents hours worked, and β represents the household's quarterly discount rate. Households obtain positive utility from consumption, which includes habit formation (Christiano et al. (2005)). Representative households earn labor income by supplying labor, but labor supply brings negative utility to the households. Households can invest in risk-free bonds B_t and deposits D_t. Deposits are assumed to be more liquid than risk-free bonds, and households gain utility by placing deposits in banks (e.g. Feenstra (1986), Drechsler et al. (2017), and Di Tella and Kurlat (2021)). This utility function implies that even when the deposit rate is lower than the rate on risk-free bonds, deposits are still an attractive investment for households.[6] The treatment of holding deposits in the utility function follows that of ABK.[7] v_t^p and v_t^d in the utility function represent demand shocks and deposit demand shocks. Each of them is determined by an autoregressive process. In particular, $v_t^p = \rho_p v_{t-1}^p + \sigma_p \epsilon_t^p, \epsilon_t^p \sim N(0,1)$ and $v_t^d = \rho_d v_{t-1}^d + \sigma_d \epsilon_t^d, \epsilon_t^d \sim N(0,1)$.

In the budget constraint equation for households, W_t denotes the nominal wage and P_t denotes the aggregate price. The inflation rate is defined as $1 + \pi_t \equiv \frac{P_t}{P_{t-1}}$. $1 + r_t^b$ stands for the nominal return on risk-free bonds. $1 + r_t^d$ is the aggregate nominal return on bank deposits. $\mathcal{P}_t$ is the aggregated profits of several sectors, including intermediate-goods-producing firms, retailers, capital good producers, banks, and financial market intermediaries, which we cover in more detail below. T_t is a lump sum transfer from the government (negative value indicates tax collection).

Deposit aggregator We assume that the deposit market is not perfectly competitive and that banks have market power. Market competition in the deposit market is assumed to take the form of

[6] In ABK, cash is also included in the model as a liquid asset. However, cash is excluded from households' portfolios in our model because, in reality, the nominal interest rate on deposits generally does not fall below zero, implying that the opportunity costs of holding bank deposits and cash are equivalent.

[7] To be precise, ABK solves the model nonlinearly. ABK uses the following formulation: $\zeta\Phi(\mathcal{L}) = -\frac{1}{2}\zeta\left(\mathcal{L}^* - \min\left\{\mathcal{L}, \mathcal{L}^*\right\}\right)^2$.

Cournot competition. The deposit aggregator applies a nested-CES function following Atkeson and Burstein (2008) as follows:

$$D_t \equiv \left[\int d_{j,t}^{\frac{\theta^d+1}{\theta^d}}\, dj\right]^{\frac{\theta^d}{\theta^d+1}}, \quad d_{j,t} \equiv \left[\sum_{i\in j} \frac{1}{\xi_{ij}^d} d_{ij,t}^{\frac{\eta^d+1}{\eta^d}}\right]^{\frac{\eta^d}{\eta^d+1}}, \quad \eta^d > 0, \quad \theta^d > 0, \tag{3}$$

where i denotes an individual bank and j denotes a market. θ^d represents across-market substitutability and captures the costs of moving across markets. Within-market substitutability, η^d, captures within-market, across-bank deposit moving costs. ξ_{ij}^d is a deposit quality adjustment shifter, the addition of which is an unusual departure from the literature. The higher the quality adjustment shifter, the lower the interest rate on deposits the bank can charge. A high quality adjustment shifter can be thought of, for example, as representing the ease of access to a bank's ATMs, its online services, or the stability of its operations from its customers' perspective. As mentioned above, this setting is necessary for making the model consistent with the stylized fact that large banks enjoy large deposit shares and lower deposit interest rates in the U.S.[8] Under this assumption, we can derive the demand for deposits as follows:

$$d_{ij,t} = \left(\xi_{ij}^d \frac{r_{ij,t}^d}{r_{j,t}^d}\right)^{\eta^d} \left(\frac{r_{j,t}^d}{r_t^d}\right)^{\theta^d} D_t \quad \Leftrightarrow \quad r_{ij,t}^d = \frac{1}{\xi_{ij}^d}\left(\frac{d_{ij,t}}{d_{j,t}}\right)^{\frac{1}{\eta^d}} \left(\frac{d_{j,t}}{D_t}\right)^{\frac{1}{\theta^d}} r_t^d. \tag{4}$$

The bank chooses the optimal deposit supply considering the inverse deposit demand curve above. See Appendix B for the full derivation.

It is useful to see the role of substitution parameters θ^d and η^d in the oligopolistic competition environment at this stage. First, the lower the degree of substitutability within markets (the smaller η^d), the greater the market power the bank has. Second, a smaller θ^d implies higher search costs for higher deposit rates across markets. In the limiting case, where $\theta^d \to 0$, local deposit market power becomes greatest because deposit transfers among markets become completely inelastic and do not respond to deposit rate differentials among markets. As across-market substitutability, θ^d, approaches infinity, representative households optimally put all of their deposits into the market with the highest deposit rates, eroding the market power of banks within the market. Third, as within-market substitutability, η^d, increases, competition intensifies as households move their deposits to banks paying higher deposit rates. Finally, as we discuss in the next section 3, under our

[8] This quality adjustment shifter is unusual in the literature that uses the Atkeson and Burstein (2008) framework. For example, Berger et al. (2022) uses the same framework as we do, but studies the labor market. The main difference between the labor market and the deposit market is the price-setting behavior of large entities. In the labor market, large firms with strong market power offer higher wages, while in the deposit market, large banks with strong market power offer lower deposit rates.

specification of preferences, elasticities and markdowns have closed-form expressions that depend only on a bank's deposit share in the market, with larger banks setting larger markdowns.

Labor unions We assume that the supply of labor by households, H_t, is determined by the demand for labor by unions. Each worker belongs to a union k. We assume that there is a continuum of unions and that each union hires a fully representative sample of workers. A competitive labor packer then packages these tasks into aggregate employment services using the following technology with elasticity of substitution that varies over time:

$$H_t = \left(\int_k H_{k,t}^{\frac{\varepsilon_t^w - 1}{\varepsilon_t^w}} dk \right)^{\frac{\varepsilon_t^w}{\varepsilon_t^w - 1}}. \tag{5}$$

The competitive labor packer then sells these services to final good firms at price W_t. We assume that there is a quadratic utility cost of adjusting the nominal wage $w_t(k)$ set by union k by including an additional additive disutility term in the household flow utility:

$$\Xi_w \left(\frac{w_t}{w_{t-1}} \right) = \int_k \frac{\kappa^w}{2} \left(\frac{w_t(k)}{w_{t-1}(k)} - 1 \right)^2 dk. \tag{6}$$

In this setup, all unions choose to set the same wage $W_t = w_t(k)$ at time t. Under this setting, we obtain the following nonlinear wage Phillips curve:

$$\pi_t^w \left(1 + \pi_t^w\right) = \frac{\varepsilon_t^w}{\kappa^w} \left(e^{v_t^p} \chi H_t^{1+\varphi} - \frac{\Lambda_t W_t^r H_t}{\mathcal{M}_t^w} \right) + \beta \pi_{t+1}^w \left(1 + \pi_{t+1}^w\right), \tag{7}$$

where $1 + \pi_t^w \equiv \frac{W_t}{W_{t-1}} = (1 + \pi_t) \frac{W_t^r}{W_{t-1}^r}$. Thus, at the zero-wage-inflation steady state, we have $\chi H^\varphi = \frac{\Lambda W^r}{\mathcal{M}^w}$.[9] Log-linearizing around the zero-wage-inflation steady state, we obtain the linear wage Phillips curve:

$$\pi_t^w = \frac{\varepsilon^w}{\kappa^w} \chi H^{1+\varphi} \left[\mu_t^w + v_t^p + \varphi h_t - \lambda_t - w_t^r \right] + \beta \mathbb{E}_t \pi_{t+1}^w. \tag{8}$$

We assume that the wage markup follows a MA(1) process, $\mu_t^w = \rho_{\mu^w} \mu^w{}_{t-1} + \epsilon_t^{\mu^w} - \theta_{\mu^w} \epsilon_{t-1}^{\mu^w}$, where $\epsilon_t^{\mu^w} \sim N(0, \sigma_{\mu^w})$.

Intermediate goods producing firms An intermediate goods firm produces goods that are sold to monopoly retailer i at a competitive nominal price P_t^{ig}. The intermediate goods firms are as-

[9] If there are no labor unions in this economy, i.e., there is no wage markup, we have $\chi H^\varphi = \Lambda W^r$ as the labor supply condition.

sumed to be established without equity capital and operate for two periods: in the first period, the intermediate-goods-producing firm borrows from the bank to rent capital and takes out a loan equal to $Q_{t-1}K_t$. K_t stands for capital and Q_t is the real price of capital in a competitive market at time t. Labor supply is sourced from the labor market at nominal wage W_t. Firms use capital and labor to operate constant returns to scale technology; in the second period, they produce and sell intermediate goods and unamortized capital to repay the debt. Thus, the problem of intermediate goods firms is:

$$\max_{K_t,H_t} \frac{P_t^{ig}}{P_t} A_t \left(K_t^{\alpha} H_t^{1-\alpha}\right)^{\nu} + (1-\delta) Q_t K_t - \frac{1+r_{t-1}^l}{1+\pi_t} Q_{t-1} K_t - \frac{W_t}{P_t} H_t, \tag{9}$$

$$A_t = A e^{v_t^a}, \quad v_t^a = \rho_a v_{t-1}^a + \sigma_a \epsilon_t^a \quad \text{where} \quad \epsilon_t^a \sim N(0,1). \tag{10}$$

We assume that total factor productivity A_t follows an autoregressive process. δ stands for the depreciation rate of capital.

Loan aggregator Intermediate-goods-producing firms raise funds L_t $(= Q_t K_{t+1})$ from the lending market at the lending rate, r_t^l. In our model, we also assume that the lending market is not perfectly competitive, so that banks have market power, and that Cournot-type quantity competition takes place. We formulate the quality adjusted non-CES aggregator as follows:

$$L_t = \left[\int l_{j,t}^{\frac{\theta^l-1}{\theta^l}} dj\right]^{\frac{\theta^l}{\theta^l-1}}, \quad l_{j,t} = \left[\sum_{i\in j} \xi_{ij}^l l_{ij,t}^{\frac{\eta^l-1}{\eta^l}}\right]^{\frac{\eta^l}{\eta^l-1}}, \quad \eta^l > 1, \quad \theta^l > 1. \tag{11}$$

The notation is identical to that used for the deposit market. However, it should be noted that the role of the quality adjustment shifter in the lending market differs from that in the deposit market. In particular, the numerator and denominator of the quality adjustment shifter are interchanged. The significance of this will be discussed below after the optimality conditions of banks are derived. Given this aggregator, solving the problem of maximizing the amount of borrowing by intermediate-goods-producing firms produces the following demand for borrowing from individual banks and relative lending rates:

$$l_{ij,t} = \left(\frac{1}{\xi_{ij}^l} \frac{r_{ij,t}^l}{r_{j,t}^l}\right)^{-\eta^l} \left(\frac{r_{j,t}^l}{r_t^l}\right)^{-\theta^l} L_t \quad \Leftrightarrow \quad r_{ij,t}^l = \xi_{ij}^l \left(\frac{l_{ij,t}}{l_{j,t}}\right)^{-\frac{1}{\eta^l}} \left(\frac{l_{j,t}}{L_t}\right)^{-\frac{1}{\theta^l}} r_t^l. \tag{12}$$

See Appendix B for the complete derivation. Based on these conditions, the bank determines the optimal loan supply.

We now describe the characteristics of the above equation, highlighting in particular the role of the quality adjustment shifter. Consider the borrowing demand as an expenditure minimization problem for intermediate-goods-producing firms. From this perspective, intermediate-goods-producing firms want to borrow from banks at as low an interest rate as possible. However, if there are additional amenities associated with banks' lending services, such as a bank's high quality of business support along with lending, firms may borrow at higher interest rates. Small and medium-sized enterprises (SMEs) are often charged higher risk premiums than larger firms because of their profit uncertainty. As a result, banks that do more business with SMEs, i.e., smaller banks, set relatively higher lending rates, as we show in Section 3. In the model, higher loan rates set by smaller banks are interpreted as representing a high quality of lending services, examples of which include frequent communication and the provision of advice often seen in relationship banking. On the other hand, from the perspective of SMEs, a bank that lends regardless of the stability of management can be viewed as a bank with high quality. From the firm's perspective, the quality adjustment shifter can be interpreted as a representation of the bank's attractiveness. The second equation above reflects these factors.

Next, we turn to the parameters for market substitutability. Basically, the role of the substitutability parameters, θ^l and η^l, is identical to that in the deposit market. First of all, the lower the degree of substitutability (i.e., the smaller θ^l or η^l), both across and within markets, the more market power a bank has. Second, a smaller θ^l implies a higher search cost for lower lending rates across markets. As θ^l becomes smaller, reallocating loans among markets becomes inelastic and less responsive to differences in lending rates among markets. As market substitutability approaches infinity, intermediate-goods-producing firms shift all their borrowing demand to the market with the lowest lending rates. Third, as is the case for within-market substitutability (η^l), as substitutability increases (as η^l increases), competition increases as intermediate-goods-producing firms move their borrowings to banks charging lower lending rates. Finally, and in contrast to the deposit market, under the above formulation we consider the markup from marginal cost in the lending market. Solving the optimization problem for banks, we get a condition that says that larger banks set higher markups.

Final goods producers A representative final good producer aggregates differentiated varieties of intermediate goods, indexed by $i \in [0, 1]$, using a CES production technology:

$$Y_t = \left(\int_0^1 y_t(i)^{\frac{\varepsilon_t^p - 1}{\varepsilon_t^p}} \mathrm{d}i \right)^{\frac{\varepsilon_t^p}{\varepsilon_t^p - 1}}, \tag{13}$$

where ε_t^p is the time-varying elasticity of substitution across differentiated intermediate input goods. Final-good-producing firms take input prices $p_t(i)$ and output prices P_t as given. The demand for inputs is given by the profit maximization condition:

$$y_t(i) = \left(\frac{p_t(i)}{P_t}\right)^{-\varepsilon_t^p} Y_t. \tag{14}$$

Under the assumptions of free entry into the final good market and zero profits in equilibrium, the price of the aggregate good is given by:

$$P_t = \left(\int_0^1 p_t(i)^{1-\varepsilon_t^p}\, \mathrm{d}i\right)^{\frac{1}{1-\varepsilon_t^p}}. \tag{15}$$

We define the gross inflation rate at time t as $\frac{P_t}{P_{t-1}} \equiv \Pi_t (= 1 + \pi_t)$. Intermediate goods firms, denoted with subscript i, produce their products with labor and capital. The quantity produced is represented by $y_t(i)$. Intermediate goods firm i chooses its optimal price $p_t(i)$ to maximize the present value of future profits:

$$\max_{p_t(i)} \sum_{t=0}^{\infty} \beta^t \Lambda_t \left[\frac{p_t(i) - P_t^{ig}}{P_t} y_t(i) - \Xi_p\left(\frac{p_t}{p_{t-1}}\right)\frac{p_t(i)}{P_t} y_t(i)\right], \tag{16}$$

where Λ_t is the real stochastic discount factor of households, who own the intermediate goods firms. $\Xi(\cdot)$ is a convex price adjustment cost function, which takes the following form: $\Xi_p\left(\frac{p_t}{p_{t-1}}\right) \equiv \frac{\kappa^p}{2}\left(\frac{p_t(i)}{p_{t-1}(i)} - 1\right)^2$. Intermediate-goods-producing firms face quadratic adjustment costs, where κ^p governs the price stickiness. The elasticity of substitution can be converted into the markup, which can be characterized by $\mathcal{M}_t \equiv \frac{\varepsilon_t^p}{\varepsilon_t^p - 1}$.

Once we log-linearize the optimality condition of the intermediate-goods-producing firms around the steady state, we obtain the following conventional New Keynesian Phillips curve:

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \frac{\varepsilon^p - 1}{\kappa^p} p_t^{ig,r} + \frac{\varepsilon^p - 1}{\kappa^p} \mu_t. \tag{17}$$

where $p_t^{ig,r}$ is the linearized expression of the real price of intermediate goods. We assume the log-deviation of the markup follows a MA(1) process, $\mu_t = \rho_\mu \mu_{t-1} + \epsilon_t^\mu - \theta_\mu \epsilon_{t-1}^\mu$, where $\epsilon_t^\mu \sim N(0, \sigma_\mu)$.

Capital goods producers The capital good producers supply capital goods to meet the quantity demanded by intermediate-goods-producing firms, whose demand is determined by the bank lending rate. In doing so, the capital good producers use the output of the final good producer as an input. The capital good producers determine the quantity of capital goods supplied so as to maximize their profit considering the real price of the capital good, Q_t. Specifically, the problem of capital good producers takes the following form:

$$\max_{I_t} \sum_{t=0}^{\infty} \beta^t \Lambda_t \left(Q_t I_t \left(1 - \Xi_i \left(\frac{I_t}{I_{t-1}} \right) \right) - I_t \right), \tag{18}$$

where $\Xi(\cdot)$ is a convex adjustment cost function for capital goods, which takes the following form: $\Xi_i\left(\frac{I_t}{I_{t-1}}\right) = \frac{\kappa^i}{2}\left(\frac{I_t}{I_{t-1}} - 1\right)^2$.

Law of motion of capital Given the decisions of capital-good-producing firms, aggregate capital evolves according to the following rule:

$$K_{t+1} = (1-\delta)K_t + I_t \left(1 - \Xi \left(\frac{I_t}{I_{t-1}} \right) \right), \tag{19}$$

where δ is the capital depreciation rate. The adjustment cost function of capital goods is identical to the adjustment cost function in the optimality problem of the capital good producers.

Monetary policy The monetary authority adjusts the interest rate of both the short-term bond and the long-term bond.[10] In the short-term bond market, the central bank follows a standard Taylor rule:

$$\frac{1+r_t^b}{1+r^b} = \left(\frac{1+r_{t-1}^b}{1+r^b} \right)^{\rho_{mp}} \left[\left(\frac{1+\pi_t}{1+\pi} \right)^{\phi^\pi} \left(\frac{Y_t}{Y} \right)^{\phi^y} \right]^{(1-\rho_{mp})} \exp\left(\sigma_{mp}\epsilon_t^{mp}\right) \quad \text{where} \quad \epsilon_t^{mp} \sim N(0,1) \tag{20}$$

Once we linearize around the steady state, we obtain the log-linearized Taylor rule as follows:

$$r_t^b = \rho_{mp} r_{t-1}^b + \left(1-\rho_{mp}\right)\left(\phi_\pi \pi_t + \phi_y y_t\right) + \sigma_{mp}\epsilon_t^{mp}. \tag{21}$$

The central bank controls long-term interest rates through quantitative easing, in which the central bank purchases long-term government bonds in the long-term government bond market. We simply assume that the long-term bond market is dominated by non-bank financial intermediaries and the central bank. This means that while the commercial banks, as described below, invest in long-term bonds, the impact of their demand on the long-term bond price is assumed to be negligible. The long-term government bond market is represented by government bonds with a maturity of 10 years. Non-bank financial intermediaries take a zero net position, shorting short-term bonds and holding long-term bonds on the asset side. 10-year government bonds are supplied inelastically by the fiscal authority, and the supply is assumed to remain unchanged. Under this assumption, the

[10] In the U.S., multiple rounds of QE were conducted as unconventional monetary policy that involved buying up large quantities of financial assets. While the level of the long-term rate was not a direct target of the policy, through QE the FED aimed to prompt investors to rebalance their portfolios in ways that lowered yields across asset classes.

yield on the 10-year treasury bond is formulated as follows:[11]

$$r_t^{10y} = \iota^{st} \cdot r_t^b - \iota^{cb} \cdot cb_t^{10y} + v_t^{mrt}. \tag{22}$$

where ι^{st} stands for the elasticity of the long-term government bond yield with respect to the short-term government bond yield, and ι^{cb} denotes the elasticity of the 10-year bond yield with respect to a long-term bond purchase shock by the central bank, cb_t^{10y}. Thus, a positive value of cb_t^{10y} indicates a quantitative easing shock, which lowers the yield of the long-term bond. We assume the log-deviation of the quantitative easing shock follows an autoregressive process, $cb_t^{10y} = \rho_{qe} cb_{t-1}^{10y} + \sigma_{qe}\epsilon_t^{qe}$, where $\epsilon_t^{qe} \sim N(0,1)$.

Fiscal policy The fiscal authority determines the size of the lump-sum tax, T_t, to satisfy the government's budget constraint, given market interest rates on short-term government debt (G_t) and long-term government debt (G^{10y}).

$$\frac{1+r_{t-1}^b}{1+\pi_t} G_{t-1} + \frac{1+r_{t-1}^{10y}}{1+\pi_t} G^{10y} + T_t \leq G_t + G^{10y}. \tag{23}$$

We assume that tax is sufficient so that households behave in a Ricardian manner and fiscal transfers do not affect the real economy. We omit the time subscript from the long-term government debt term as we assume that its size remains constant throughout time in this economy.

2.2 Banking system block

In this section, we formulate the problem that each bank faces and solve the optimality condition for the problem. We assume that the banks' objective is to maximize a flow dividend for each period. This implies that each bank maximizes the net interest inome by solving a one-period problem.

Banks make corporate loans $l_{i,t}$ and invest in long-term government bonds $b_{i,t}^l$ using their net worth $n_{i,t}$ as well as deposits $d_{i,t}$. The law of motion of net worth of each bank is characterized by:

$$n_{i,t+1} = (1-\gamma)\left(n_{i,t} + nii_{i,t}\right), \tag{24}$$

where $n_{i,t}$ is the net worth of bank i, $nii_{i,t}$ is the net interest income of bank i, and parameter $\gamma \in \left(\beta^{-1} - 1, 1\right)$ determines the dividend share paid to households.[12]

[11] This formulation is based on the market segmentation hypothesis (e.g. Vayanos and Vila (2021) and Gourinchas et al. (2022)) and demand asset pricing (e.g. Koijen and Yogo (2019) and Gabaix and Koijen (2021)). See Appendix A for derivation. For those who are interested in the derivation for a setting with non-bank financial intermediaries, Abe (2025) is useful.

[12] It should be noted that the banks' problem is to maximize one-period interest income in our model. Other HBANK models, such as that of Bellifemine et al. (2022), deal with the banks' problem by solving the value function using

To maximize the net worth at time $t+1$, each bank i optimally chooses its supply of deposits, $d_{i,t}$, loan volume, $l_{i,t}$, and duration of assets, $\psi_{i,t}$. The problem of bank i is:

$$\arg \max_{d_{i,t}, l_{i,t}, \psi_{i,t}} [n_{i,t+1}] = \arg \max_{d_{i,t}, l_{i,t}, \psi_{i,t}} [(1-\gamma)(n_{i,t} + nii_{i,t})] = \arg \max_{d_{i,t}, l_{i,t}, \psi_{i,t}} [nii_{i,t}]. \tag{25}$$

The problem of banks corresponds to maximizing one-period net interest income, since the net worth of the bank is a state variable. In the following, we consider how a bank's choice of deposits, loan volume and asset duration affect net interest income, but we temporarily ignore the term related to the inflation rate to simplify the argument. We come back to this point in the last part of this section. Note that the asset side of the bank's balance sheet is composed of loans and investments in long-term government bonds. On the liability side, banks have deposits and net worth. Thus, the net interest income for bank i is:

$$\arg \max_{d_{i,t}, l_{i,t}, \psi_{i,t}} [nii_{i,t}] = \arg \max_{d_{i,t}, l_{i,t}, \psi_{i,t}} \left[r^{bl}_{i,t} b^{l}_{i,t} + r^{l}_{i,t} l_{i,t} - r^{d}_{i,t} d_{i,t} - \left(mnnie^{l}_{i} \cdot l_{i,t} - mnnii^{d}_{i} \cdot d_{i,t} \right) \right], \tag{26}$$

where $b^{l}_{i,t}$ is the amount of security investment, $d_{i,t}$ and $l_{i,t}$ are the amounts of deposits and loans of bank i, $mnnie^{l}_{i}$ is the marginal net non-interest expense in the lending market and $mnnii^{d}_{i}$ is the marginal net non-interest income from the deposit market, which we estimate in the sprit of Bellifemine et al. (2022).[13] Bank i sets deposit rate $r^{d}_{i,t}$ and lending rate $r^{l}_{i,t}$ in incomplete markets, which we describe below. Finally, $r^{bl}_{i,t}$ represents the blended return of the bank's portfolio of loans and security investment in longer-term government bonds. The long-term rate set by bank i is determined by duration, $\psi_{i,t}$, in the following manner:

$$r^{bl}_{i,t} = \psi_{i,t} r^{10y}_{t} + (1-\psi_{i,t}) r^{b}_{t} = \psi_{i,t} \left(r^{10y}_{t} - r^{b}_{t} \right) + r^{b}_{t} = \psi_{i,t} r^{sp}_{t} + r^{b}_{t}. \tag{27}$$

Thus, the choice variables of bank i are $d_{i,t}, l_{i,t}$, and $\psi_{i,t}$. We divide the bank's problem into two stages. In the first stage, banks compete in both deposit and loan markets to set the optimal supply of deposits and loans. In the second stage, banks choose the optimal duration for assets given the size of assets and the excess return and volatility of 10-year bonds. In other words, we assume that the optimal duration for assets, $\psi_{i,t}$, is already determined by the first stage, since the duration is chosen in the second stage of time period $t-1$. The following equations formulate the problem of banks in stage 1 and stage 2.

backward iteration and dealing with the distribution using a forward equation, as in Auclert et al. (2021). However, the outcome of our model is close to that of Bellifemine et al. (2022) in the sense that the response to a shock, such as monetary shock, of the HBANK model is larger than that of the RBANK model.

[13]The details of this estimation are in Appendix B.

Stage 1: competition in deposit and loan markets We formulate the banks' problem at each time period as follows:

$$
\begin{aligned}
nii_{i,t} = & \max_{d_{i,t}, l_{i,t}, \psi_{i,t}} \left[r^{bl}_{i,t} b^{l}_{i,t} + r^{l}_{i,t} l_{i,t} - r^{d}_{i,t} d_{i,t} - \left(mnnie^{l}_{i} \cdot l_{i,t} - mnnii^{d}_{i} \cdot d_{i,t} \right) \right] \\
& \text{s.t.} \quad b^{l}_{i,t} + l_{i,t} \leq n_{i,t} + d_{i,t},
\end{aligned}
\tag{28}
$$

where we add a balance sheet budget constraint. Rewriting the nii maximization problem of bank i with the balance sheet constraint yields

$$
\max_{d_{i,t}, l_{i,t}, \psi_{i,t}} \left[\underbrace{\left(r^{b}_{t} + mnnii^{d}_{i} - r^{d}_{i,t} \right) d_{i,t}}_{\text{deposit market}} + \underbrace{\left(r^{l}_{i,t} - r^{bl}_{i,t} - mnnie^{l}_{i} \right) l_{i,t}}_{\text{lending market}} + \left(r^{bl}_{i,t} - r^{b}_{t} \right) \left(d_{i,t} + n_{i,t} \right) + r^{b}_{t} n_{i,t} \right]. \tag{29}
$$

We assume that the duration of the base rate for the deposit market is the same as that of the three-month short-term rate. Thus, banks earn additional profit by taking duration risk. We can separate the problems in the deposit market and the lending market into the following problems:

$$
\text{deposit market:} \quad d^{*}_{i,t} = \arg\max_{d_{i,t}} \left[\left(r^{b}_{t} + mnnii^{d}_{i} \right) d_{i,t} - r^{d}_{i,t} \left(d_{i,t}, d^{*}_{-i,t}, r^{d}_{t}, D_{t} \right) d_{i,t} \right], \tag{30}
$$

$$
\text{lending market:} \quad l^{*}_{i,t} = \arg\max_{l_{i,t}} \left[r^{l}_{i,t} \left(l_{i,t}, l^{*}_{-i,t}, r^{l}_{t}, L_{t} \right) l_{i,t} - \left(r^{bl}_{i,t} + mnnie^{l}_{i} \right) l_{i,t} \right], \tag{31}
$$

where $*$ denotes the optimal deposit supply and loan supply. Given the relative demand from (4) and (12), each bank solves for the optimal supply of deposits and loans. Again, note that the optimal duration for the first stage problem is already determined in the previous period.

Stage 2: choice of duration for asset side As documented in Drechsler et al. (2021), a bank's asset duration is key to understanding the cross-sectional heterogeneity in income β^{Inc} among banks. Thus, we incorporate banks' choice of asset duration, $\psi_{i,t}$. A bank's asset duration matters for the return and volatility of the asset side of its balance sheet. Each bank seeks a higher yield, but wants to avoid the greater volatility that can be generated from taking on longer duration. Therefore, following He and Krishnamurthy (2013), we assume that banks take the following mean-variance strategy, accounting for the quadratic term:[14]

$$\max \psi_{i,t} \mathbb{E}_{t-} \left[\psi_{i,t} r_t^{sp} \mathcal{A}_{i,t} \right] - \frac{\Gamma_{i,t}}{2} \mathbb{V}_{t-} \left[\psi_{i,t} r_t^{sp} \mathcal{A}_{i,t} \right] \tag{32}$$

where $\mathcal{A}_{i,t}$ is the asset size of bank i at time t and is composed of loans and security investments. $\Gamma_{i,t}$ is a time-variant parameter that controls the risk aversion of the bank. Thus, the optimal asset duration for bank i is

$$\psi_{i,t} = \frac{1}{\Gamma_{i,t} \mathcal{A}_{i,t}} \frac{\mathbb{E}_{t-} \left[r_t^{sp} \right]}{\Sigma}, \tag{33}$$

where we assume that the volatility of the long-term bond is time-invariant, i.e., $\Sigma \equiv \mathbb{V}_{t-} \left[r_t^{sp} \right]$. The optimality condition for duration implies that banks prolong asset duration as the expected return on long-term assets relative to safe assets increases. On the other hand, banks shorten their asset duration as they become more risk averse, their total asset size increases, or the volatility of long-term assets increases. We assume that the volatility of the return on long-term assets is time-invariant in the following analysis.

The key assumption is that we impose duration risk in both loans and security investments. Drechsler et al. (2021) documents that larger banks tend to have lower repricing maturity and a higher share of short-term assets on their balance sheets.[15] To explain the negative correlation between asset size and asset duration in our model, we refer to Drechsler et al. (2021), which argues that this cross-sectional heterogeneity arises from the composition effect between loans and securities, where the duration of securities tends to be longer than that of loans.

Finally, we assume the risk-taking parameter follows an autoregressive process:

$$\Gamma_{i,t} = \Gamma_i \exp \left(-v_t^{rt} \right), \tag{34}$$

where $v_t^{rt} = \rho_{rt} v_{t-1}^{rt} + \sigma_{rt} \epsilon_t^{rt}, \epsilon_t^{rt} \sim N(0,1)$ and "rt" indicates *risk taking*. Given a shock that increases v_t^{rt}, *all* banks take more duration risk simultaneously.

Net interest income with inflation In the above argument, we disregard the impact of the inflation rate on the return of each asset. If we consider inflation in the economy, then the law of motion

[14] Specifically, He and Krishnamurthy (2013) employs a mean-variance strategy to determine how much of their balance sheets households allocate to risky assets through their investment in financial intermediaries' equity, but the general idea that an agent chooses what fraction of his assets to allocate to risky and safe assets using the mean-variance strategy is identical.

[15] The share of short-term assets refers to loans and security investments with repricing maturity of less than one year as a percentage of total assets.

of the real net worth and interest income of bank i becomes:

$$n_{i,t} + nii_{i,t} = \frac{1 + r_t^{bl}}{1 + \pi_{t+1}} n_{i,t} + \frac{r_t^l - r_t^{bl}}{1 + \pi_{t+1}} l_{i,t} + \frac{r_t^{bl} - r_t^d}{1 + \pi_{t+1}} d_{i,t}. \tag{35}$$

2.3 Equilibrium definition

Thus far we have described the equilibrium system of the NK block and the heterogenous bank sector. We wrap up this section by defining the general equilibrium system of the economy.

Equilibrium system *Given stochastic processes for the deposit demand shock, the preference shock, the productivity shock, the markup shock, the conventional monetary policy shock, the quantitative easing shock, the banks' risk-taking shock, initial price level* P_{-1}*, government short-term debt* G_{-1}*, capital stock* K*, and the initial distribution of the net worth of banks* $\mathcal{D}_0(n_i)$*, a competitive equilibrium is a stochastic sequence of prices* $\{P_t, W_t, P_t^{ig}, Q_t, r_t^b, r_t^{10y}, r_{i,t}^d, r_{i,t}^l\}$*, aggregates* $\{C_t, H_t, B_t, D_t, \Pi_t^F, \Pi_t^B, T_t, K_t, Y_t, I_t\}$*, individual policy rules* $\{d_{i,t}, l_{i,t}, \psi_{i,t}\}$*, and joint distribution of banks such that:*

(i) $\{C_t, H_t, B_t, D_t\}$ *maximizes the infinite horizon utility subject to the budget constraint;*

(ii) Decisions $\{H_t, K_t\}$ *solve the intermediate-goods-producing firms' problem, taking* $\{P_t^{ig}, P_t, r_t^l, Q_t, W_t, A_t\}$ *as given, and ensuring that loan demand* L_t *is consistent with the solution to the intermediate firms' problem;*

(iii) Prices $\{p_t(i)\}_{i\in[0,1]}$ *and intermediate goods demand* $\{y_t(i)\}_{i\in[0,1]}$ *solve the monopolistically competitive final good producers' problem, taking prices* P_t^{ig}, P_t *and the stochastic discount factor* Λ_t *as given;*

(iv) Π_t^F *is equal to the aggregate profits of monopolistic final good producers and capital producers given their decisions;*

(v) Investment I_t *solves the problem of capital good producers, taking the price of capital* Q_t *and the stochastic discount factor* Λ_t *as given;*

(vi) Decisions $\{d_{i,t}, l_{i,t}, \psi_{i,t}\}$ *solve the individual bank's problem, taking* $\{r_t^b, r_t^{10y}, D_t, L_t, P_t\}$ *as given, individual bank net worth* $n_{i,t}$ *follows the law of motion given by (24), and* Π_t^B *is equal to aggregate bank dividends given these decisions;*

(vii) Monetary policy is set according to (20) and transfers T_t *are set to satisfy the government's intertemporal budget constraint; and*

(viii) the market clearing conditions for the final good, labor, the capital good, short-term bonds, long-term bonds, loans, and deposits are satisfied.

2.4 Solution method

Our solution method builds on the sequence-space Jacobian method developed in Auclert et al. (2021). When deriving the steady state in our equilibrium system, we use investment and the aggregate deposit rate as unknown variables, i.e., $\mathbf{U} = \left(I, r^{d,1}\right)$. We compute both the macro NK block and the banking system block to find the $\mathbf{U}$ that solves $\mathbf{H}(\mathbf{U}) = 0$. $\mathbf{H}(\mathbf{U})$ is a market clearing condition in which we include the lending rate and the deposit rate$\left(r^{l,1} - r^{l,2}, r^{d,1} - r^{d,2}\right)$. To do this, we use the following iterative procedure. First, starting from $j = 1$, we guess values for $\mathbf{U}^1$. Second, we calculate $\mathbf{H}\left(\mathbf{U}^1\right)$. Third, we form a guess for $\mathbf{U}^{j+1}$ using the following procedure:

$$\mathbf{U}^{j+1} = \mathbf{U}^j - \left[\mathbf{H}_{\mathbf{U}}\left(\mathbf{U}^j\right)\right]^{-1}\mathbf{H}\left(\mathbf{U}^j\right). \tag{36}$$

Once $\mathbf{H}(\mathbf{U})$ converges to 0, we stop the iteration. Note that the initial input for $r^{d,1}$ and output from the equilibrium system $r^{d,2}$ are different. One of the unknowns, $r^{d,1}$, is used to derive the demand for deposits, D, in the macro NK block. The other unknown, I, is used to compute the demand for capital, K. Given D and K, the banking system block generates the lending rate and deposit rate, $\left(r^{l,2}, r^{d,2}\right)$. We take the difference between $\left(r^{l,2}, r^{d,2}\right)$ generated in the banking system block and $\left(r^{l,1}, r^{d,1}\right)$ generated in the NK block or imported as an unknown. If these two values converge, we stop the iteration.

3 Micro Calibration for Banking System

In this section, we show the optimality conditions of banks for pricing in the deposit and loan markets and outline their characteristics. Then, we explain our calibration strategy. In the calibration section, we first show how we construct the target datasets and then describe the strategy for calibrating the model using a general equilibrium model that combines the NK block and the banking system block. In the latter part, we present the calibration results and explore their implications. In the results section, we compare our heterogeneous bank model with the representative bank model, which assumes that individual banks are homogeneous and representative in terms of their price setting behavior.

3.1 Optimal price setting behaviors of banks

In our model, the banks have market power over the deposit and loan markets. As we show in the pervious section, bank solves the optimal problems for supply of deposits and loans, (30) and (31) given the relative demand indicated by (4) and (12). The solutions to those two problems can be formulated as follows.[16] We start with the optimal condition in the deposit market.

Optimality condition in the deposit market. The optimal deposit rate set by bank i at time t is:

$$\text{deposit rate}: \quad r^d_{i,t} = \mu^d_{i,t} mr^d_{i,t}, \tag{37}$$

[16] A full derivation for the optimality conditions is in Appendix B.

where $\mu^d_{i,t}$ denotes the markdown. $mr^d_{i,t}$ stands for the marginal revenue of supplying an additional unit of deposits. They take the following forms:

$$\begin{aligned} \text{markdown}: \quad & \mu^d_{i,t} = \frac{1}{1 + \left[\boldsymbol{s^d_{i,t}} \frac{1}{\theta^d} + \left(1 - \boldsymbol{s^d_{i,t}}\right) \frac{1}{\eta^d} \right]}, \\ \text{marginal revenue}: \quad & mr^d_{i,t} \equiv r^b_t + mnnii^d_i, \end{aligned} \tag{38}$$

where $\boldsymbol{s^d_{i,t}}$ denotes the deposit sales share and $s^d_{i,t}$ stands for the share of outstanding deposits, which can be formulated mathematically as follows:

$$\begin{aligned} \text{deposit sales share}: \quad & \boldsymbol{s^d_{i,t}} \equiv \frac{r^d_{i,t} d_{i,t}}{r^d_t D_t} = \zeta^{d\,\eta^d}_i \left(\frac{r^d_{i,t}}{r^d_t} \right)^{1+\eta^d}, \\ \text{deposit share}: \quad & s^d_{i,t} \equiv \frac{d_{i,t}}{D_t} = \zeta^{d\,\eta^d}_i \left(\frac{r^d_{i,t}}{r^d_t} \right)^{\eta^d}. \end{aligned} \tag{39}$$

$mnnii^d_i$ is the marginal net non-interest income of bank i. Before explaining the intuition of the formulas above, we should note an important assumption on market structure. We assume that the deposit market is symmetric across markets j (i.e., across regions). Mathematically speaking, we assume that $r^d_{j,t} = r^d_t$. This assumption allows us to identify the key substitution parameters without region-specific deposit rates.

The marginal revenue denotes the return that a bank earns by obtaining one additional unit of deposits. In our model, banks can invest their deposits at short-term interest rates without duration risk. Banks with deposit outflow risk, i.e., liquidity constraints, obtain an additional liquidity premium on the return from deposits, which is represented by $mnnii^d_i$.[17] We assume that the deposit rate is this marginal revenue multiplied by its markdown.

The markdown is determined by the inter- and within-market substitutability parameters, θ^d and η^d, and the bank's deposit sales share. The size of the markdown varies across banks and comes from the bank's deposit sales share.[18] It should be noted that the deposit sales share and share of deposits outstanding differ only in the relative price of each bank, but qualitatively they have a similar implication for the market power of banks in the deposit market. Another unique aspect is the quality adjustment shifter. The larger the quality adjustment shifter, the larger the sales share

[17]Section B.3 shows that $mnnnii^d_t$ can be constructed by adding a liquidity constraint to the bank's problem.

[18]In the literature that uses the Atkeson and Burstein (2008) framework, it is often assumed that $\eta > \theta$. This implies that entities with larger market shares can set lower markdowns. However, the literature that uses this oligopolistic competition framework for banks is limited and, to the best of the authors' knowledge, there is no concensus on the relationship between banks' deposit market share and their markdowns. For this reason, the restriction that $\eta > \theta$ is not included in our calibration procedure.

of deposits. Therefore, it is possible that banks with lower deposit rates may have a larger share of deposits than banks with higher deposit rates.

At this stage it is worthwhile mentioning a source of heterogeneity of deposit services among banks in our model. The deposit markdown and deposit share are endogeneous variables determined by market competition. Banks are intrinsically heterogeneous with respect to their service quality (ξ_i^d) and marginal net non-interest income ($mnnii_i^d$), which are estimated.

Optimal condition in the loan market We derive the optimality conditions for the loan market in an analogous way to the deposit market. The optimal lending rate set by bank i is:

$$\text{lending rate}: \quad r_{i,t}^l = \mu_{i,t}^l mc_{i,t}^l, \tag{40}$$

where $\mu_{i,t}^l$ is the loan markup and $mc_{i,t}^l$ is the marginal cost incurred to generate one unit of loans. These two quantities take the following forms:

$$\begin{aligned} \text{markup}: \quad & \mu_{i,t}^l = \frac{1}{1 - \left[\boldsymbol{s_{i,t}^l} \frac{1}{\theta^l} + \left(1 - \boldsymbol{s_{i,t}^l}\right) \frac{1}{\eta^l}\right]}, \\ \text{marginal cost}: \quad & mc_{i,t}^l \equiv r_{i,t}^{bl} + mnnie_i^l. \end{aligned} \tag{41}$$

We define the loan sales share and the share of loans outstanding as follows:

$$\begin{aligned} \text{loan sales share}: \quad & \boldsymbol{s_{i,t}^l} \equiv \frac{r_{i,t}^l l_{i,t}}{r_t^l L_t} = \xi_i^{l\eta_l} \left(\frac{r_{i,t}^l}{r_t^l}\right)^{1-\eta^l}, \\ \text{loan share}: \quad & s_{i,t}^l \equiv \frac{l_{i,t}}{L_t} = \xi_i^{l\eta_l} \left(\frac{r_{i,t}^l}{r_t^l}\right)^{-\eta^l}. \end{aligned} \tag{42}$$

We assume that differences in the pricing behavior across regions are symmetric, as in the deposit market. The base interest rate included in the marginal cost, $r_{i,t}^{bl}$, takes into account the duration risk of banks.

As we will discuss below, the marginal net non-interest expense refers to the net cost of a bank's lending business. We assume that this is captured by $mnnie_i^l$. In other words, we consider $mnnie_i^l$ as capturing the cost of monitoring business conditions of firms. As we discuss in Appendix B.4, the estimated $mnnie_i^l$ for smaller banks is larger.

On the other hand, from the perspective of firms, a bank that continues to lend regardless of the business conditions of firms can be seen as a high-quality bank. The quality adjustment shifter in our model can be thought of as reflecting the attractiveness of bank lending from the perspective of firms. As can be seen in the formulation of the market share, the higher the quality of the bank, the lower the relative lending rate and the larger the market share. Banks with a high market share are able to charge higher markups.

3.2 Data

We use Call Report data for individual banks in the U.S., as in Drechsler et al. (2017).[19] The period of analysis is 35 years, from 1985 to 2020. The sample includes 171 non-community banks, whose assets combined represent about 70% of the total assets of all U.S. banks, including community banks. According to the FED, a bank with total assets of less than $10 billion is defined as a community bank.[20] However, this definition continues to change as the total asset size of banks changes. For this reason, the sample here includes non-community banks that have been at least once in the top 3% of banks in terms of total assets in the sample period.

Estimation of $mnnii_i^d$ and $mnnie_i^l$ In preparation for the main calibration, we follow the methodology developed in Bellifemine et al. (2022) and Jamilov and Monacelli (2023) to estimate marginal net non-interest income ($mnnii_i^d$) in the deposit market and marginal net non-interest expenses ($mnnie_i^l$) in the loan market.

Loan duration The data for duration risk in lending is calculated by weighting the maturity of each bank's loans. Banks report their holdings of loans and leases broken down into six bins by repricing maturity interval (0 to 3 months, 3 to 12 months, 1 to 3 years, 3 to 5 years, 5 to 15 years, and over 15 years). To calculate the repricing maturity, we assign the interval midpoint to each bin (and 15 years to the last bin) and take a weighted average using the amounts in each bin as weights.

Subsamples To see how the competition environment for banks has changed over time, we divide the full sample period into three subsamples, from 1984 to 2008, from 2008 to 2017, and from 2017 to 2020. The subsample period starting in 2008 corresponds to the period of zero nominal interest rates. Through the lens of our model, we see how the zero lower bound period affects our interpretation of competition among banks.

When conducting the following calibration, we take an average of each of the following variables within the target period: $\{r_i^d\}$, $\{mr_i^d\}$, $\{s_i^d\}$, $\{r_i^l\}$, $\{mc_i^l\}$, $\{s_i^l\}$, $\{\psi_i\}$. More details on the datasets and estimation procedure are provided in Appendix B.

3.3 Calibration strategy

Here we describe our calibration strategy for the across- and within-market substitutability parameters and the quality adjustment shifter in both deposit and lending markets, adding to the risk-taking parameters. In Section 2.4, we describe how we solve our equilibrium system, which incorporates

[19] We use the datasets available on Philipp Schnabl's website.

[20] https://www.federalreserve.gov/supervisionreg/community-and-regional-financial-institutions.htm.

both the NK block and the banking system block. To calibrate the substitution parameters and the quality adjustment shifters, we combine the following outer and inner loops.

Outer 0. Guess an initial set of deposit and lending rates, r^d and r^l.

Outer 1. Given the deposit and lending rates, compute demand for loans and deposits, K and D, using the NK macroeconomic block.

Inner 1. Deposit market: Given $\{r_i^d\}$, $\{s_i^d\}$ and $\{mr_i^d\}$, we use a grid search to estimate the θ^d, η^d, and $\{\xi_i^d\}$ that minimize the following loss function, which is given by the Euclidean distance between individual bank data and the model. We describe the grid search that we use in the calibration to find the best set of θ^d and η^d in the next paragraph.

$$\min_{\theta^d,\eta^d,\{\xi_i^d\}} \|\widehat{r_i^d} - r_i^d\| \tag{43}$$

$$\widehat{\xi_i^d} = \frac{r_t^d}{r_{i,t}^d} s_{i,t}^{d\ \frac{1}{\eta^d}} \tag{44}$$

Inner 2. Loan market: Given $\{r_i^l\}$, $\{s_i^l\}$ and $\{mc_i^l\}$, we use a grid search to estimate the θ^l, η^l, and $\{\xi_i^l\}$ that minimize the loss function, which is given by the Euclidean distance between individual bank data and the model.

$$\min_{\theta^l,\eta^l,\{\xi_i^l\}} \|\widehat{r_i^l} - r_i^l\| \tag{45}$$

$$\widehat{\xi_i^l} = \frac{r_{i,t}^l}{r_t^l} s_{i,t}^{l\ \frac{1}{\eta^l}} \tag{46}$$

Inner 3. Generate the model-implied deposit rates and shares, $\{\widehat{r_i^d}\}$ and $\{\widehat{s_i^d}\}$, using $\{mr_i^d\}$, $\widehat{\theta^d}$, $\widehat{\eta^d}$ and $\{\widehat{\xi_i^d}\}$. For the loan market, generate the model-implied lending rates and shares, $\{\widehat{r_i^l}\}$ and $\{\widehat{s_i^l}\}$, in an analogous way.

Inner 4. Since K and D are given from the macro block, using the model-implied deposit rate, deposit share, lending rate and lending share, the individual bank's net interest income is determined by:

$$\widehat{n}_i = \frac{1-\gamma}{\gamma}\widehat{nii}_i \tag{47}$$

The dividend rule of the bank and the interest income of individual banks determine the steady-state net assets.

$$\widehat{\mathcal{A}_i} = \widehat{s_i^d} D + \widehat{n}_i \tag{48}$$

Outer 2. Compute the individual risk aversion parameter from the estimated balance sheet size, the long-term bond spread and the data for individual banks' duration risk.

$$\widehat{\mathbf{\Gamma_i}} = \frac{r^{sp}}{\psi_i \widehat{\mathcal{A}_i}} \tag{49}$$

Then, return to **inner loop step 1**. Once the model-implied net worth of the bank converges, we stop the iteration.

Grid search In the **inner loop step 1** and **inner loop step 2** described above, we estimate the parameters for within- and across-market substitution using the deposit and loan interest rates suggested by the model and the data. We use a grid search for this calibration. Specifically, when calibrating the parameters for the deposit market, we examine all possible combinations of η^d (within-market substitutability) and θ^d (across-market substitutability) within a two-dimensional log grid with 40,000 grid points in the interval [0.01, 100].

Given a single pair of η^d and θ^d, we use the following equation to calculate the individual bank's quality adjustment shifter:

$$\widehat{\xi_i^d} = \frac{r_t^d}{r_{i,t}^d} s_{i,t}^d{}^{\frac{1}{\widehat{\eta^d}}}.$$

In this calculation, we use data for the individual bank's deposit interest rate, the macro deposit interest rate, and the individual bank's market share together with the calibrated $\widehat{\eta^d}$. Given $\widehat{\eta^d}$ and $\widehat{\theta^d}$, and knowing $\widehat{\xi_i^d}$, the implied markdown and deposit rate in the model can be calculated using eqn. (41) and the marginal return on deposits. Finally, we calculate the loss function as the difference between the model-implied deposit interest rate and the data. We apply this calculation to all grid points and use the $\widehat{\eta^d}$, $\widehat{\theta^d}$ and $\widehat{\xi_i^d}$ with the smallest loss function as the estimated results.

The same method is applied to the loan market. However, in the loan market, the parameters for the within-market and across-market substitution are greater than 1, so the range of grid points is taken to be [1, 10000]. Finally, it is important to note that in both the deposit and loan markets, there are no constraints on the within- and across- market substitution parameters.

Calibrated parameters in the NK block When performing the calibration procedure above, we need the macro NK block to calculate the demand for capital (K) and deposits (D) that are consistent with the proposed deposit and loan rates. Moreover, capital and deposits are also required to determine the net worth and the size of the balance sheet of banks in the banking system block. Therefore, it is necessary to determine the parameters to be calibrated in both the macro NK block and the banking system block. We describe them in Table 1.

On the one hand, most of the parameters determining the steady state follow the values in the literature. On the other hand, some parameters are calibrated according to the approach of ABK in

Table 1: *Calibrated* parameters for the steady state

Parameter		Value	Description (literature)
Households			
Discount factor	β	0.9975	Real short term rate: 1.0% (p.a.)
Relative risk aversion	σ	1	Christiano et al. (2005), Christiano et al. (2014)
Habit formation	h	0.62	Christiano et al. (2005), Smets and Wouters (2007)
Disutility of labor	χ	0.41	Abadi, Brunnermeier, and Koby (2023)
Frisch elasticity of labor supply	φ	2	Chetty et al. (2011)
Wage markup	$\mathcal{M}^w$	1.1	Auclert et al. (2021)
Deposit demand	ζ	0.0045	Loan-to-bond ratio, $\frac{L}{B^L}$
Deposit satiation point	D^*	14	Deposit-GDP ratio, $\frac{D}{Y}$
Intermediate-goods-producing firms			
Capital share	α	0.36	Christiano et al. (2005), Gertler and Karadi (2011)
Scale parameter	ν	0.8	Consumption-investment ratio
Capital depreciation rate	δ	0.025	Christiano et al. (2005), Smets and Wouters (2007)
Firm productivity	A^*	1.0	Normalization
Final-good-producing firms			
Retail price elasticity	ε^p	3.9	Gertler and Karadi (2011), $\mathcal{M} = 1.34$
Banking system block			
Bank dividend payout rate	γ	0.06	Capitalization ratio, $\frac{N}{L}$
Long-term yield	r^{10y}	0.0125	Nominal 10 year term rate: 5.0% (p.a.)

order to be consistent with the macro data. The scale parameter ν of the production function is set to the ratio of consumption to investment during the period of analysis. The liquid asset saturation point, D^*, which is crucial for the steady-state level of deposits, is determined by the ratio of deposits to real GDP. γ, which determines the dividend rule of the bank, is set so that the capitalization ratio, N/L, of the bank is consistent with the data. We only consider the dynamics of consumption and corporate investment and do not account for the government sector or net exports. Thus when calibrating the model we use the average consumption and investment to GDP ratio during the sample period, $(C + I)/Y = 0.836$.

3.4 Calibration results

Next, we present the results of our calibration. In addition to the results for the entire sample period, we also show the results using subsample periods. In analyzing these subsample periods, we discuss changes in the competition environment along with the level of market interest rates, characteristics of the estimation in the zero lower bound period, and characteristics of the model.

Table 2: Estimated parameters for the deposit market and the loan market

Parameter	84Q2-08Q1	08Q2-17Q1	17Q2-20Q3	84Q2-20Q3
Deposit within-market substitution, η^d	1.57	0.87	0.27	1.42
Deposit across-market substitution, θ^d	0.11	0.02	0.04	0.09
Loan within-market substitution, η^l	7.64	2.80	3.78	5.73
Loan across-market substitution, θ^l	5.17	1.35	2.44	3.70

3.4.1 Deposit market

First, we investigate the deposit market. The calibrated substitution parameters are summarized in Table 2. The results of the estimation for the deposit market are shown in Figure 2. The top panels show the deposit interest rates for each bank and the marginal return on deposits for each analysis period, as well as the deposit interest rates implied by the model based on the calibrated results. The middle panels show the markdowns implied by the model. The bottom panels show the estimated quality adjustment shifters.

Looking at the results for the whole sample period, we see that large banks tend to have small markdowns (large deposit spreads). This finding is in line with Jamilov and Monacelli (2023). This is primarily due to the fact that the level of deposit interest rates is low and the marginal return on deposits is large for larger banks. From the bottom panel, we see that larger banks exhibit high quality in their deposit operations. The level of deposit interest rates can be kept low due to the high quality of large banks. Again, it is important to note that when calibrating, we do not impose constraints on the size of the parameters for the within- and across-market substitution.

There are two points to be noted in the results of the subsample analysis. The first is that the markdown rate decreases (deposit spread increases) for the two subsample periods after the zero interest rate policy began compared to the first subsample period. One potential explanation behind this could be that demand for deposits from households increased. If demand for deposits increases, banks increase their market power in the deposit market and charge smaller markdowns.

The second point is that, in the third subsample period, there is little heterogeneity in the markdown rates among banks. This is due to the fact that, as the market interest rate rises from the zero lower bound, the relative contribution of net non-interest income from deposit services decreases. As a result, the heterogeneity in the marginal return on deposits becomes smaller and there are no longer any differences in markdowns. Thus, competition to acquire deposits may have become more intense in this period. Moreover, heterogeneity in the quality adjustment shifter increased. This is consistent with Figure 1 in showing that concentration of the deposit market has increased.

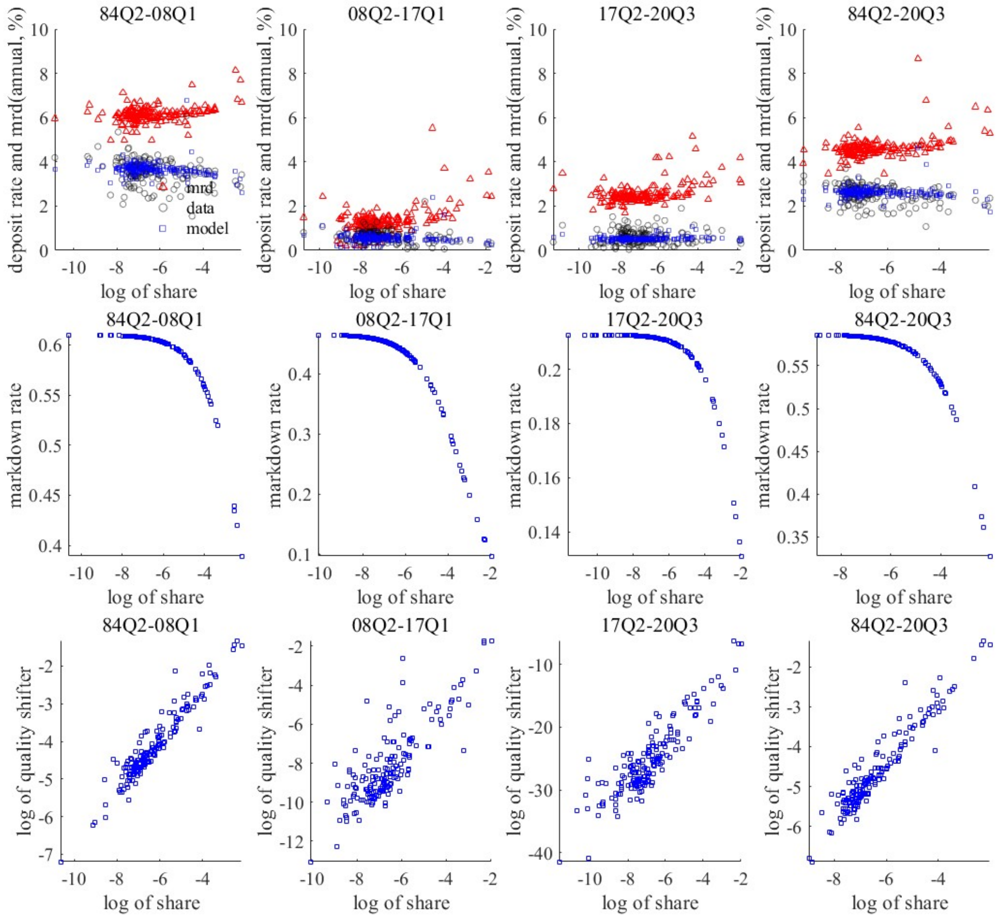

Figure 2: Comparison of deposit rates between data and model and estimated marginal revenue from deposits, deposit markdown computed in the model, quality adjustment shifters of deposit operations in the U.S.

3.4.2 Loan market

The results of the calibration for the loan market are summarized in Table 2 and Figure 2. The structure of the panel is the same as that of the deposit market. First, the large banks charge higher markups for all subsample periods. This finding is consistent with Corbae and D'Erasmo (2021) and Jamilov and Monacelli (2023). In addition, the larger banks have higher loan quality and, as a result, they capture a larger share of relative demand than they would have based only on their relatively low lending rates.

Next, let us look at the analysis using the subsamples. When comparing the first subsample period before the introduction of QE with the second and third subsample periods after the introduction of QE, we see that the markup rate increases after the introduction of QE. In the model, this is due to the fact that the marginal cost of lending for banks decreased due to the decline in long-term interest rates caused by QE. One can also see that the markup rate decreases from the second to the third subsample period. This is because, although market interest rates rose, banks kept the interest rates on their fixed-rate loans constant and the changes in lending rates remained

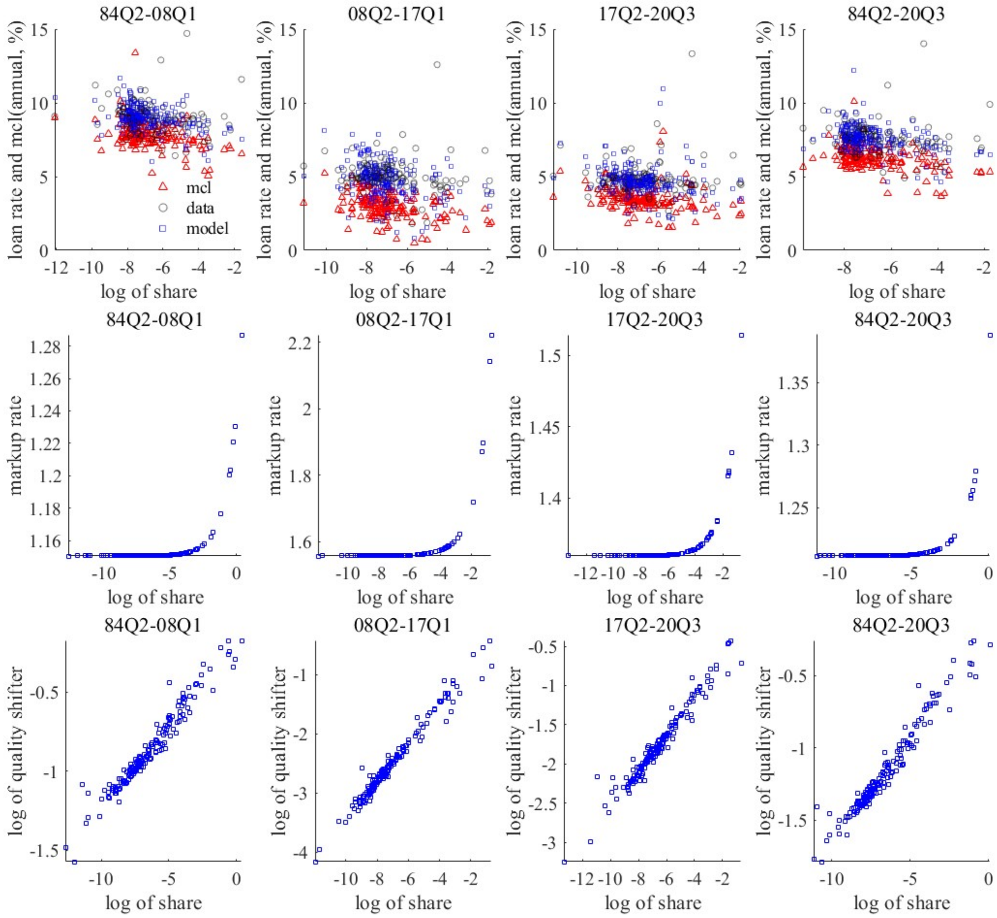

Figure 3: Comparison of loan rate between data and model, and estimated marginal cost for loan, loan markup computed in the model, quality shifters of lending operation in the U.S.

small. Finally, in comparing the first and third subsamples, one can see that the heterogeneity in the loan markup between small banks and large banks widened. After the zero lower bound period ended, large banks were able to raise lending rates as market interest rates rose while small banks were unable to do so.

3.4.3 Comprehensive comparison of data and model

In the following section we perform a dynamic model analysis with exogenous shocks. In the analysis, heterogeneity among banks in terms of the shares of their balance sheets composed of deposits and net worth is key to understanding banks' duration risk. For this reason, the model used in the analysis should be consistent with the data regarding banks' interest rate setting behavior and their market shares in the deposit and lending markets. Figure 4 shows the interest rates set by individual banks in the deposit and loan markets as well as the market share of individual banks in the data and the model. Figure 4 confirms that the model is generally consistent with the data, both qualitatively and quantitatively. At the steady state, large banks that set their lending rates low also set their deposit rates low. In addition, banks with a large share of the lending market also have a large share

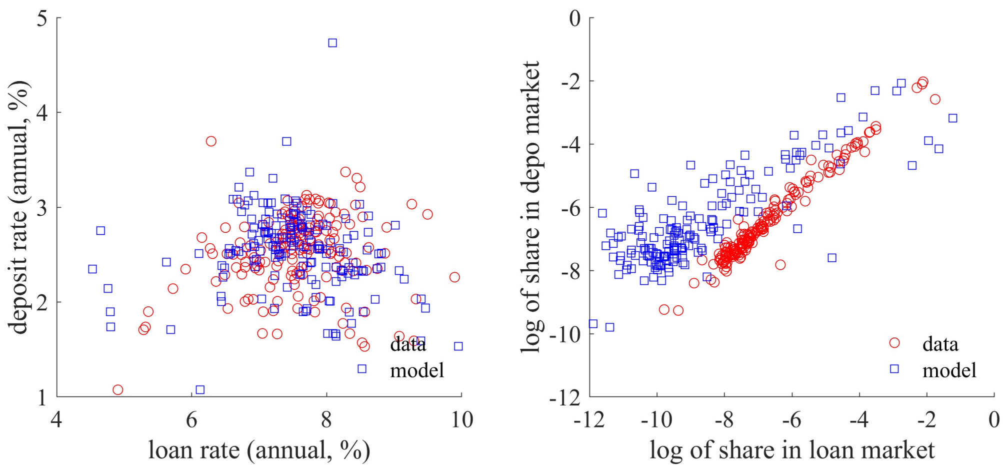

Figure 4: Model-data comparison of price setting and market share

of the deposit market.

Second, we relate our model to the findings in Drechsler et al. (2021). Drechsler et al. (2021) finds that the pass-through rates of interest rate increases to both the asset side (β_i^{Inc}) and the liability side (β_i^{Exp}) are higher for large banks. Our model is consistent with theirs with respect to the asset side because the estimated higher markup and shorter duration imply that β_i^{Inc} is higher for larger banks. However, the lower deposit markdown for larger banks is not in line with the higher β_i^{Exp} in Drechsler et al. (2021). This is mainly due to the fact that their β_i^{Exp} is pushed up by the presence of wholesale funding, which has an expense beta of one, while we do not account for any liabilities other than deposits.

3.5 Introduction of a representative bank NK model

In the following sections, we compare the HBANK model with a representative bank NK model, which assumes that individual banks are homogeneous but have market power in the deposit and loan markets. By comparing these two models, we show that the presence of heterogeneity is important when analyzing the transmission mechanism of exogenous shocks such as monetary policy shocks and productivity shocks.

When analyzing the differences in dynamics around the steady state between the HBANK model and the RBANK model, it is desirable that aggregate prices in the HBANK model and the RBANK model be identical at the steady state. To this end, we make the following adjustments when calibrating the parameters for within-market and across-market substitution in the deposit and loan markets for the RBANK model. Marginal net non-interest income in the deposit market, marginal net non-interest expenses in the loan market, the duration risk parameter, and the quality adjustment shifter in each market are calculated as the average of the values calibrated for the HBANK model weighted by the individual bank's share in each market. Using these weighted averages, we cali-

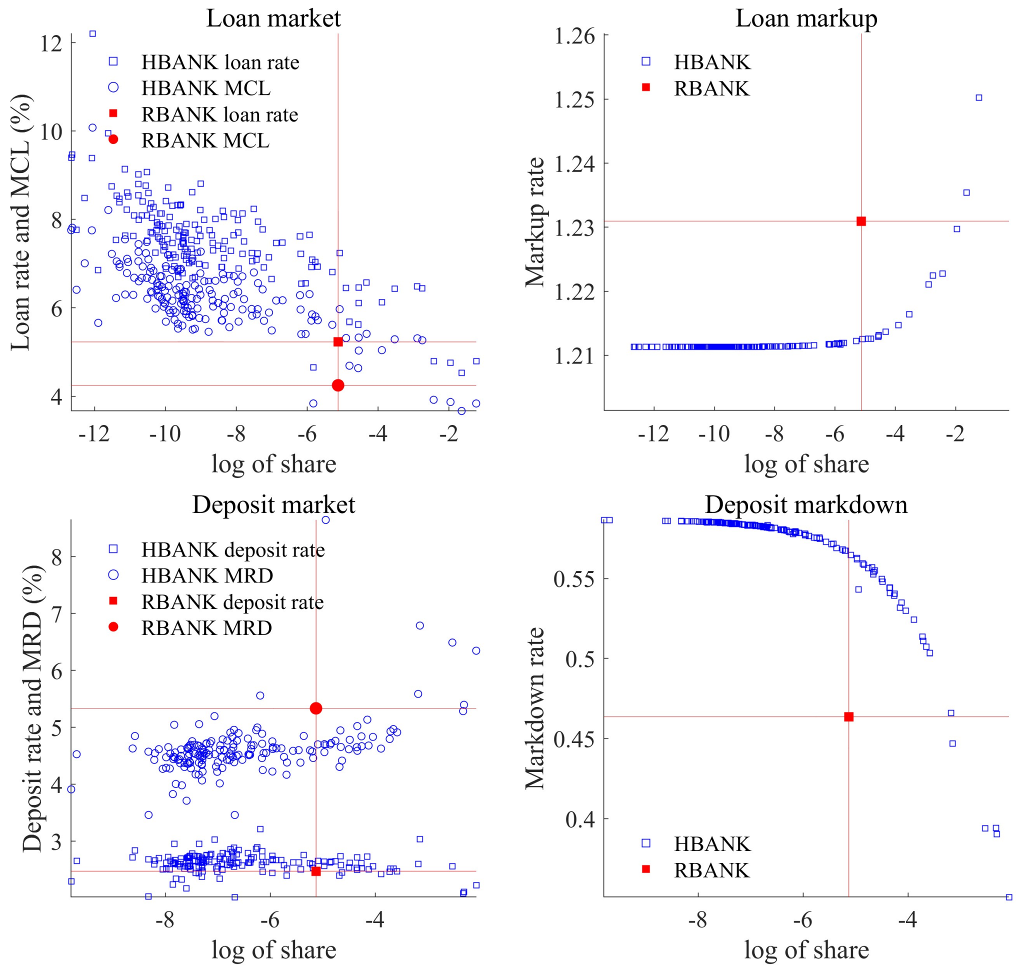

Figure 5: Comparison of loan market and deposit market of HBANK model with RBANK model at the steady state

brate the parameters for within- and across- market substitution. Figure 5 shows the distributions of markups and markdowns in the deposit and loan markets at the steady state for the HBANK and the RBANK models. The macro-level markup and markdown are identical between the two models. In addition, the duration risk parameter for the RBANK model is calibrated so that it is consistent with the duration of loans weighted by the share of loans outstanding.

4 Macro Implication of HBANK

In this section, we compare the HBANK model, a heterogeneous bank model that assumes oligopolistic competition in the deposit and loan markets, with the RBANK model, a model in which the behavior of banks is homogeneous. This allows us to analyze the effect of competition among banks on monetary policy and other macro shocks, and to highlight the characteristics of the HBANK model. In this section we analyze short-term interest rate hike shocks and positive productivity shocks, but the basic insights do not change for other shocks. Note that the macro variables of the HBANK and

Table 3: *Calibrated* parameters for dynamic analysis

Description	Parameter	Value	Description (literature)
Monetary policy response to inflation	ϕ_π	1.2	Auclert et al. (2021)
Monetary policy response to output gap	ϕ_y	0.2	Auclert et al. (2021)
Elasticity of 10-year bond to QE	ι^{cb}	0.0015	Chung et al. (2023)
Capital adjustment cost	κ^i	5.0	Abadi et al. (2023)
Wage adjustment cost	κ^w	1.0	Auclert et al. (2024)
Price adjustment cost	κ^p	70	Abadi et al. (2023)
Elasticity of LT bonds to ST bonds	ι^{st}	0.0034	*Estimated with HBANK model*

the RBANK models are calibrated so that allocations and prices of these models are identical at the steady state. After investigating the mechanism of the HBANK model, we check for consistency of the heterogeneity of bank behavior with past data.

When computing the dynamic behavior around the steady state, we use the parameters in Table 3. We calibrate the monetary policy parameters for responses to both inflation and output. Specifically, we use the values in Auclert et al. (2021), which are conventional values in the NK literature. We follow Chung et al. (2023) in calibrating the elasticity of the 10-year bond yield with respect to the FED's asset purchase program. In particular, we calibrate the parameter so that *"the median effects are consistent with an expansion of Treasury securities holdings by about $2.7 trillion and a decline in term premium effects of approximately 80 basis points"* (Chung et al. (2023)). For the adjustment cost parameters for capital, wages, and prices, we use the values commonly found in the literature. Finally, we use the parameters estimated in Section 5 for the elasticity of long-term bonds with respect to short-term bonds.

4.1 Monetary policy shock

4.1.1 IRFs analysis

Figure 6 shows the IRFs in response to a rise in short-term interest rates. The response of corporate investment in the HBANK model is larger than in the RBANK model, while the reponses of short- and long- term yields are similar. The larger reponse of corporate investment in the HBANK model is due to the large increase in the loan rate in the model compared with the RABNK model. In the following, we decompose the factors affecting these results using a partial equilibrium model focusing on the banking system block.

4.1.2 Partial equilibrium analysis

Why does the loan rate increase more in the HBANK model than in the RBANK model? Here, we analyze the mechanism by decomposing the response of the banking system block to a positive

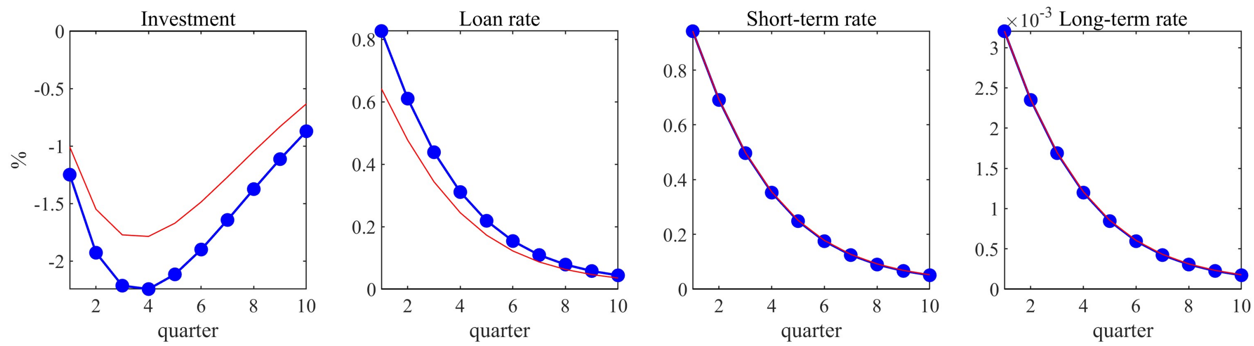

Figure 6: IRFs to 1% monetary policy hike shock
NOTE: We show the IRFs for the case where the AR (1) parameter of the monetary shock is $\rho_{mp} = 0.5$. Blue solid circle lines represent IRFs of the HBANK model and red solid lines represent IRFs of the RBANK model.

monetary policy shock. Before moving on to the partial equilibrium results, we describe the key factors to understand the dynamics of aggregate lending and deposit rates. The basic idea is that the aggregate lending and deposit rates are determined by the interest rates set by individual banks and the market share of these banks. Thus, we can decompose the aggregate lending and deposit rate responses into three terms - two first order terms and a second order term.

$$dr_t^l = \sum_i \left[s_{i,t+1}^l r_{t+1}^l - s_{i,t}^l r_t^l\right] = \sum_i \left[s_{i,t}^l \underbrace{\Delta r_t^l}_{\text{loan rate chg.}} + \underbrace{\Delta s_{i,t}^l}_{\text{loan shift}} r_t^l + \mathcal{O}\left(\Delta^2\right)\right] \tag{50}$$

$$dr_t^d = \sum_i \left[s_{i,t+1}^d r_{t+1}^d - s_{i,t}^d r_t^d\right] = \sum_i \left[s_{i,t}^d \underbrace{\Delta r_t^d}_{\text{dep. rate chg.}} + \underbrace{\Delta s_{i,t}^d}_{\text{dep. shift}} r_t^d + \mathcal{O}\left(\Delta^2\right)\right] \tag{51}$$

We focus on the first order terms - changes in interest rates and changes in volumes outstanding for both loan and deposit markets - in the following discussion.

Duration risk First, Figure 7 looks at changes in the duration risk channel that affect the marginal cost of lending operations. The risk-taking channel depends on the difference between long- and short-term interest rates in the financial market and the size of the bank's balance sheet. Here, the size of the balance sheet is determined by the bank's net worth and the size of its deposits, but in a partial equilibrium analysis, there is no significant change in the size of deposits of individual banks. Therefore, we narrow our discussion down to the interest rate differential between long- and short-term interest rates, which has a greater impact. When short-term interest rates rise, the degree of risk-taking by banks shrinks, regardless of their size. The important point here is that the risk-taking attitude of banks at the steady state determines the degree of pass-through of short-term interest rates to the marginal cost of lending. More specifically, banks with short maturities are more likely to reflect increases in short-term interest rates in the marginal cost of lending. On the other hand,

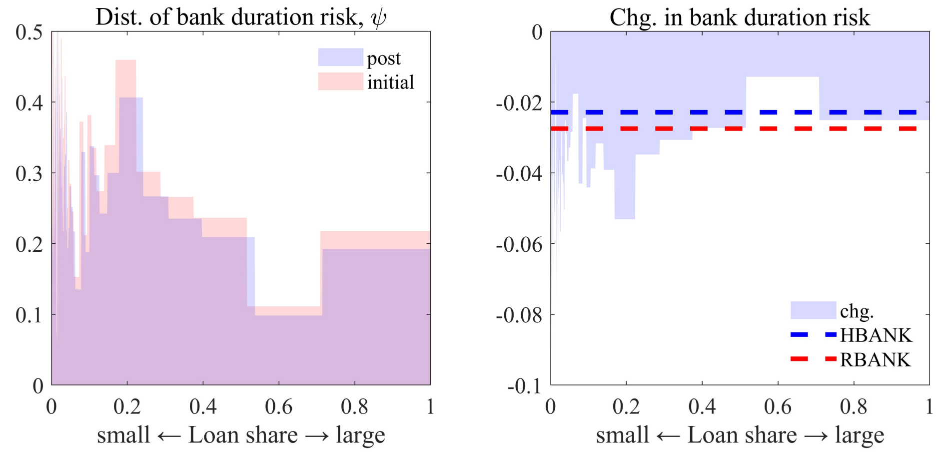

Figure 7: PE response to 1% policy rate hike

NOTE: On the left panel, the horizontal axis shows the share of each bank at the steady state and in the post-shock period.

banks with long duration reflect short-term interest rate increases in marginal costs with a lag.

Loan market Figure 8 shows the changes in lending rates in response to an increase in short-term market interest rates. The increase in lending rates is greater for large banks, indicating that the pass-through of lending rates to the market short-term rate is greater for larger banks. Two different mechanisms play a role in explaining the heterogeneity in the degree of loan pass-through. First, as argued in the previous paragraph, larger banks with short maturities are more likely to reflect increases in short-term interest rates in the marginal cost of lending.

Second, the different sizes of loan markups among banks also play an important role. Remember that the lending rate is determined by the change in the marginal cost of lending and the loan markup. The loan markup is larger for major banks, as we discuss in Section 3. Thus, if changes in the marginal cost of lending are uniform across banks, then the increase in lending rates would be larger for major banks. In other words, incorporating banks' risk-taking behavior with respect to duration enhances the pass-through of lending rates more for large banks.

As a result of the increase in lending rates for large banks, their share of the lending market decreases. Looking at the macro pass-through rate, the HBANK model has a higher pass-through rate than the RBANK model because firms shift loans to smaller banks with higher lending rates at the steady state in the HBANK model. Admittedly, this mechanism contradicts the findings of Kashyap and Stein (2000), where smaller banks contract lending more in response to a monetary tightning shock than larger banks. However, in the real economy, monetary tightning often occurs when the economy is growing and productivity is improving. As we argue in the next section, a positive productivity shock shifts demand from smaller banks to larger banks, which is consistent with the findings of Kashyap and Stein (2000).

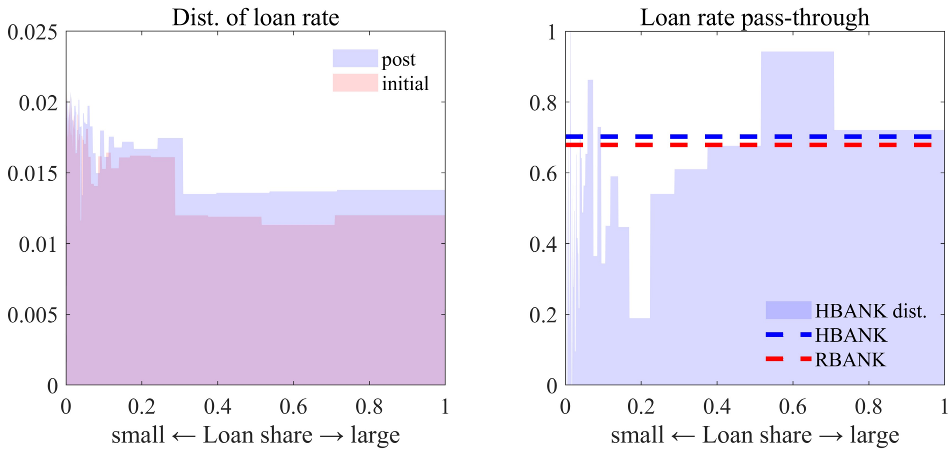

Figure 8: PE responses to 1% policy rate hike
NOTE: On the left panel, the horizontal axis shows the share of each bank at the steady state and in the post-shock period.

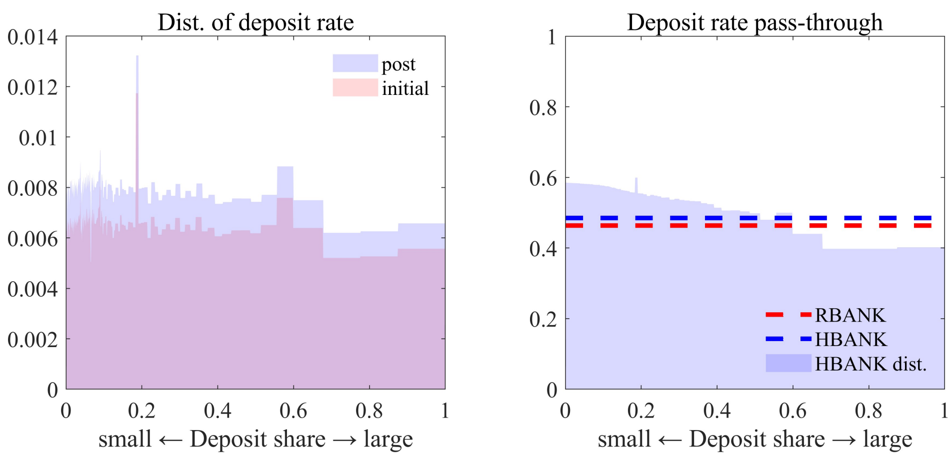

Figure 9: PE response to 1% policy rate hike
NOTE: On the left panel, the horizontal axis shows the share of each bank at the steady state and in the post-shock period.

Deposit market Figure 9 shows the changes in deposit rates in response to changes in short-term interest rates. The increase in deposit rates is larger for smaller banks, and the pass-through rate to short-term interest rates is larger for smaller banks. The market share of large banks in the deposit market declines, as in the loan market, because the increase in deposit rates is smaller. The pass-through rate at the macro level is slightly higher in the HBANK model than in the RBANK model, as deposits shift to smaller banks with higher deposit rates at the steady state.

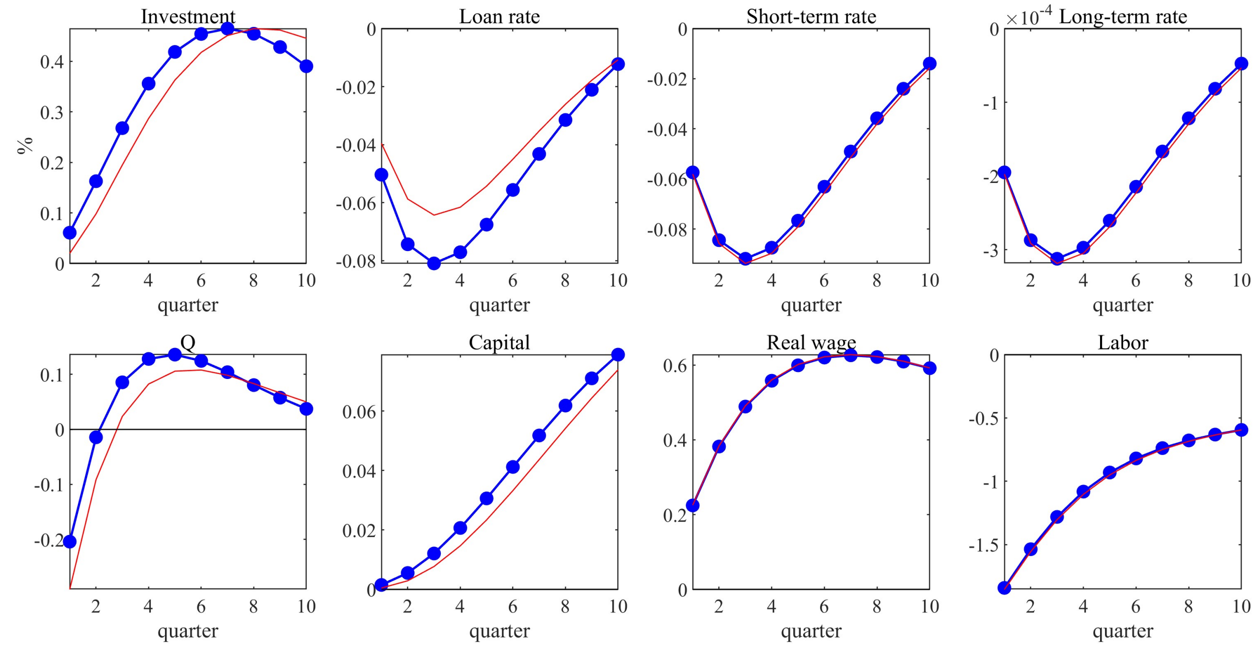

Figure 10: IRFs to 1% positive productivity shock

NOTE: We show the IRFs for the case where the AR (1) parameter of the productivity shock is $\rho_a = 0.9$. Blue solid circle lines represent the IRFs of the HBANK model and red solid lines represent the IRFs of the RBANK model.

4.2 Productivity shock

4.2.1 IRF analysis

Figure 10 shows the IRFs in response to a positive 1% TFP shock. The long- and short-term market interest rates fall to the same extent in both the HBANK and the RBANK models. However, the response of lending rates is larger in the HBANK model, as in the case of a monetary shock.

4.2.2 Partial equilibrium analysis

The right panel of Figure 11 shows the results of a partial equilibrium analysis in which a positive productivity shock is added to the banking system block. The pass-through rate of the lending rate to the market rate for large banks is relatively large, and this is due to the fact that the marginal cost of lending of large banks decreases significantly due to their short duration. As a result, the macro lending rate falls due to the loan shift effect, in which demand for loans shifts to major banks with low lending rates at the steady state. This is the main mechanism through which a positive productivity shock causes lending rates to fall more in the HBANK model.

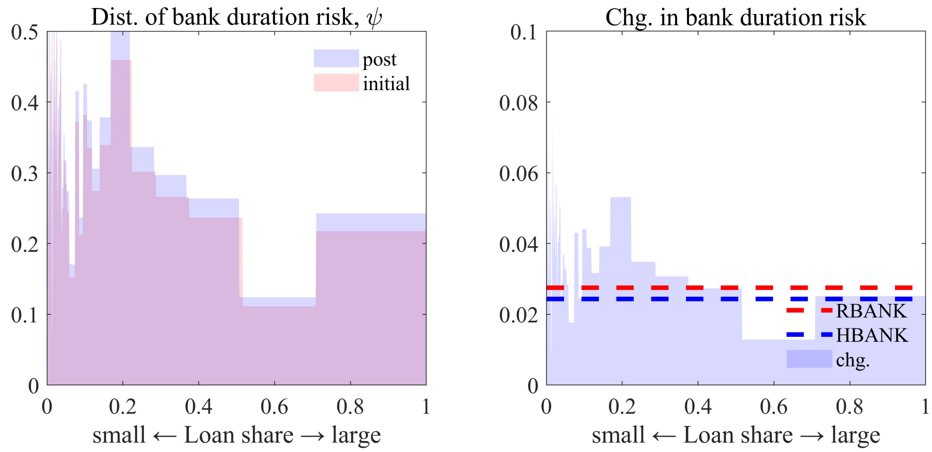

Figure 11: PE response to 1% positve productivity shock

Note: On the left panel, the horizontal axis shows the share of each bank at the steady state and in the post-shock period.

5 Macro Estimation

In the previous section, we qualitatively analyzed the differences in the transmission channels of exogenous shocks to market interest rates and productivity between the HBANK and the RBANK models. In this section, we perform Bayesian estimation using macro time series data to examine the quantitative implications of introducing heterogeneity and competition into the banking system.

5.1 Data and Methodology

Estimation target In order to focus on the differences between the HBANK and the RBANK models, the parameters we target with Bayesian estimation are shock process parameters. Specifically, we estimate the AR and MA parameters, as well as the standard deviations, of all shocks. In other words, we calibrate most of the structural parameters, such as those related to households' habits and the production function using values borrowed from the literature. The four exceptions are the three adjustment cost parameters for nominal prices, nominal wages, and capital investment, and the elasticity of the long-term interest rate with respect to the short-term interest rate. These four parameters are estimated with the HBANK model and fed into the estimation of the RBANK model.

The following nine shocks are estimated: monetary policy shock to the short-term interest rate, QE shock to the long-term interest rate, banks' duration risk shock, household deposit demand shock, risk-taking shock in the long-term bond market, markup shock in the final good market, wage markup shock in the labor market, productivity shock, and demand shock. We assume that all these shocks are orthogonal to each other.

Data The data used for the estimation comprise nine series: GDP, consumption, the inflation rate, the nominal wage inflation rate, the FF rate, the 10-year U.S.Treasury bond interest rate, the Fed's holdings of U.S.Treasury bonds, the lending rate, and deposits outstanding (y, c, π, π^w, r_t^b, r_t^{10y}, cb, r_t^l, d). The data period is the same as for the micro-moment analysis - 35 years (140 quarters), from 1985 to 2020. We summarize the details of the datasets in Appendix C.2.1.

For GDP and consumption, we use the level after detrending using the HP filter. For inflation, nominal wage inflation, the FF rate, and the 10-year U.S.Treasury rate, we demean by subtracting the period average. For the Fed's holdings of U.S.Treasuries, we use the change in the level from the 2003 level, when data first became available, as the effect of the Fed's purchases of government bonds. For deposits, we divide by the trend component of the GDP series to create a series relative to GDP, and then use the series after demeaning it by the period average.

Methodology The typical approach to likelihood-based estimation in the DSGE literature is to compute the likelihood by applying the Kalman filter to the model's state-space representation (e.g. Smets and Wouters (2007) and Herbst and Schorfheide (2016)). However, since we employ a MA representation with the sequence-space Jacobian method, as in the HANK literature (e.g. Auclert et al. (2021)), we use the Whittle approximation to compute the likelihood, as in Hansen and Sargent (1981) and Plagborg-Møller (2019). This efficient way to calculate the likelihood using the Fast Fourier Transform enables us to rapidly compute the likelihood.[21]

5.2 Estimation results

5.2.1 Duration risk shock

Figure 12 shows the historical decomposition of the output gap.[22] The first remarkable takeaway is the large role of the duration risk shock. In the post-GFC era, we observe that the duration risk shock largely offsets the positive impact of the long- and the short-term interest rate on changes in output. This is due to the fact that the loan rate does not drop as much as the long- and the short-term interest rates after the GFC. This is similar to the finding in Gerali et al. (2010) that the positive loan markup shock muted the effect of QE by the ECB in the Euro area.

Figure 13 shows the forecast error variance decompositions for both the HBANK and the RBANK models. The main takeaway from this exercise is that the bank's risk-taking shock plays a large role in explaining the variation of output for both the HBANK and the RBANK models. As discussed above, in our model, duration risk dampens the responses of short-term interest rates to output. Since the loan rate remained largely at the same level as the pre-GFC period, the duration risk shock explains the lion's share of corporate investment during the zero lower bound period. Comparing the HBANK and the RBANK models, the duration risk channel has almost the same impact as the bank

[21]Recent studies attempt to incorporate cross-sectional information to perform Bayesian estimation of heterogeneous-agent models (e.g. Chang et al. (2021) and Liu and Plagborg-Møller (2023)). However, we do not pursue this strategy here.

[22]The estimated parameters are summarized in Table C.2 and the historical decomposition for the other observable variables are described in Figure C.2.

credit channel on corporate investment. Moreover, the size of these risk-taking shocks is comparable to that found in related studies (Gerali et al. (2010)). This suggests that the endogenous risk-taking channel in our model is limited.

5.2.2 Productivity shock and markup shock

As we saw in the previous section, the response of corporate investment to a TFP shock is larger in the HBANK model due to the duration risk channel and credit shift channel. The estimated HBANK model shows that the variance due to the TFP shock is three times as large as in the estimated RBANK model. Instead, we find that the contribution of markup shocks is smaller in the HBANK model than in the RBANK model. This is possibly due to differences in the dynamics of lending rates between the HBANK and the RBANK models, as well as differences in the nature of productivity and markup shocks on the inputs of the production function. Specifically, a positive productivity shock leads to an increase in the real wage, which in turn causes a decline in prices (Figure 10). The short-term interest rate and the loan rate drop due to a fall in prices and the conventional dynamics of the Taylor rule. As a result, the intermediate-goods-producing firms shift their demand for inputs from labor to capital. A decrease in labor due to a positive productivity shock features as a *technology shock* channel in Gali (1999).

On the other hand, a negative markup shock lowers prices, inducing a decline in the short-term interest rate and hence an increase in aggregate demand.[23] A decline in prices causes the real wage to rise due to nominal wage stickiness. Simultaneously, a decline in prices induces a drop in the lending rate and hence an increase in corporate investment and capital. However, the increase in the real wage and the demand for capital are smaller than in the case of a positive productivity shock.

Thus, the only difference between these two shocks comes from the behavior of labor in equilibrium. In the case of a positive productivity shock, where the real wage rises, a decline in the lending rate induces more corporate investment due to the input shift channel from labor to capital. On the other hand, in the case of a negative markup shock, where the increase in the real wage is modest, the lending channel of corporate investment is muted relative to a productivity shock. In the HBANK model, a decline in the lending rate is amplified, thus enhancing the productivity shock channel from labor to capital. Therefore, the impact of productivity shocks on corporate investment and hence output is greater in the HBANK model.

[23]Figure C.1 shows the IRFs to a negative price markup shock.

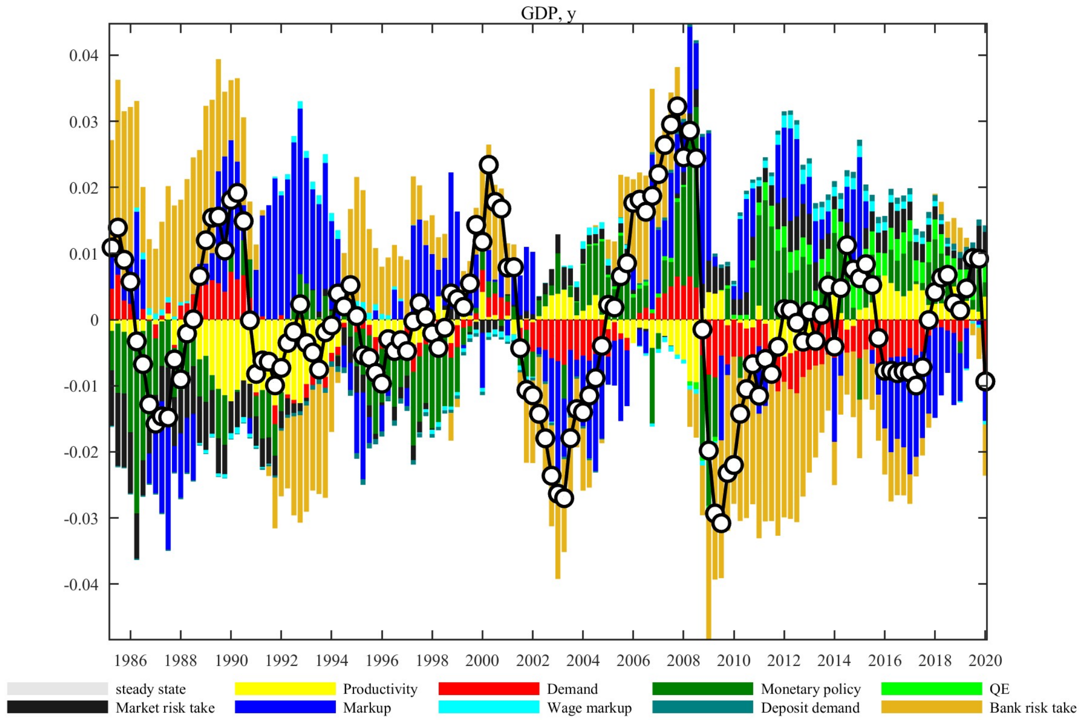

Figure 12: Historical decomposition of the output gap

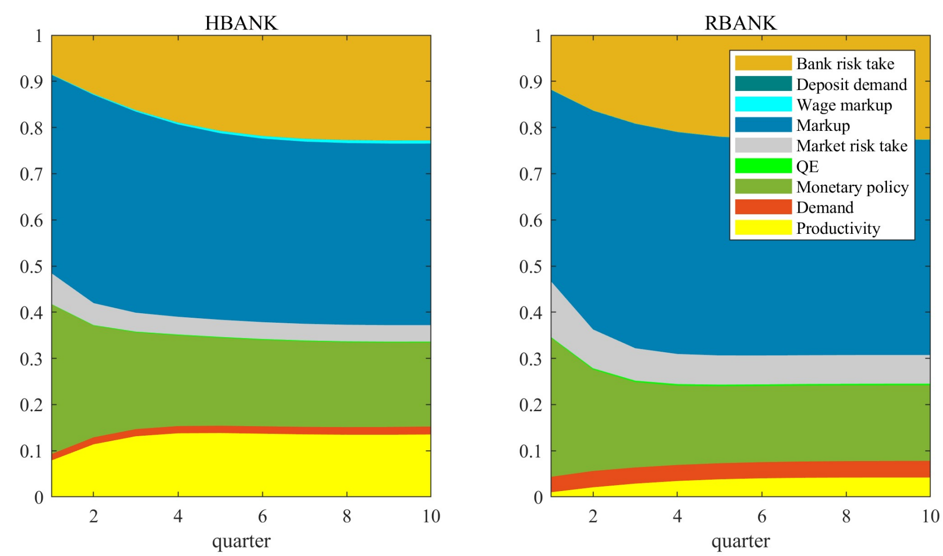

Figure 13: Variance decomposition of GDP, left panel: HBANK, right panel: RBANK

Note: This figure decomposes $\mathrm{Var}_{t-1}(Y_{t+h})$ into contributions of the nine shocks of the model.

6 Conclusion

In this paper, we develop a HBANK model, which incorporates oligopolistic competition among banks in the deposit and loan markets, to analyze the macro implications of bank heterogeneity and competition. We calibrate the banking sector with U.S.non-community banks using the Call Report data. In our model, the individual level pass-through (β) of deposit rates and loan rates can be made consistent with the data by incorporating a quality adjustment shifter and loan-duration-risk behavior. Using the HBANK model, we analyze how the transmission mechanisms of macroeconomic shocks change when heterogeneity among banks is taken into account. We find that the bank lending channel is stronger in the HBANK model than the RBANK model. In other words, our HBANK model amplifies macroeconomic shocks, such as monetary policy shocks and productivity shocks. An important lesson is that the effects of the duration risk channel and credit shifting channel among banks are large. Finally, we analyze how consideration of heterogeneity changes our understanding of past business cycles. Using Bayesian estimation with time series datasets, we find that the contribution of productivity shocks to GDP is about three times as large in the HBANK model as in the RBANK model.

References

Abadi, Joseph, Markus Brunnermeier, and Yann Koby. 2023. "The reversal interest rate." *American Economic Review* 113 (8): 2084–2120.

Abe, Nobuhiro. 2025. "FTPL Redux with Market Segmentation." Mimeo.

Amiti, Mary, Oleg Itskhoki, and Jozef Konings. 2019. "International shocks, variable markups, and domestic prices." *The Review of Economic Studies* 86 (6): 2356–2402.

Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing-to-market, trade costs, and international relative prices." *American Economic Review* 98 (5): 1998–2031.

Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub. 2021. "Using the sequence-space Jacobian to solve and estimate heterogeneous-agent models." *Econometrica* 89 (5): 2375–2408.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2020. "Micro jumps, macro humps: Monetary policy and business cycles in an estimated HANK model."Technical report, National Bureau of Economic Research.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2024. "The intertemporal keynesian cross." *Journal of Political Economy* 132 (12): 4068–4121.

Bellifemine, Marco, Rustam Jamilov, and Tommaso Monacelli. 2022. "Hbank: Monetary policy with heterogeneous banks."

Berger, David, Kyle Herkenhoff, and Simon Mongey. 2022. "Labor market power." *American Economic Review* 112 (4): 1147–1193.

Bernanke, Ben S, and Alan S Blinder. 1988. "Credit, money, and aggregate demand."

Bernanke, Ben S, and Mark Gertler. 1995. "Inside the black box: the credit channel of monetary policy transmission." *Journal of Economic perspectives* 9 (4): 27–48.

Bianchi, Javier, and Saki Bigio. 2022. "Banks, liquidity management, and monetary policy." *Econometrica* 90 (1): 391–454.

Boppart, Timo, Per Krusell, and Kurt Mitman. 2018. "Exploiting MIT shocks in heterogeneous-agent economies: the impulse response as a numerical derivative." *Journal of Economic Dynamics and Control* 89 68–92.

Boyd, John H, and Gianni De Nicolo. 2005. "The theory of bank risk taking and competition revisited." *The Journal of finance* 60 (3): 1329–1343.

Brunnermeier, Markus K, and Yuliy Sannikov. 2014. "A macroeconomic model with a financial sector." *American Economic Review* 104 (2): 379–421.

Buch, Claudia M, Sandra Eickmeier, and Esteban Prieto. 2014. "In search for yield? Survey-based evidence on bank risk taking." *Journal of Economic Dynamics and Control* 43 12–30.

Campbell, John Y, and Luis M Viceira. 2002. *Strategic asset allocation: portfolio choice for long-term investors.* Clarendon Lectures in Economic.

Chang, Minsu, Xiaohong Chen, and Frank Schorfheide. 2021. "Heterogeneity and aggregate fluctuations."

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2011. "Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins." *American Economic Review* 101 (3): 471–475.

Christiano, Lawrence J, Martin Eichenbaum, and Charles L Evans. 2005. “Nominal rigidities and the dynamic effects of a shock to monetary policy.” *Journal of political Economy* 113 (1): 1–45.

Christiano, Lawrence J, Roberto Motto, and Massimo Rostagno. 2014. “Risk shocks.” *American Economic Review* 104 (1): 27–65.

Chung, Hess, Etienne Gagnon, James Hebden, Kyungmin Kim, Bernd Schlusche, Eric Till, James Trevino, and Diego Vilán. 2023. “Balance Sheet Policies in an Evolving Economy: Some Modelling Advances and Illustrative Simulations.”

Coimbra, Nuno, and Hélene Rey. 2024. “Financial cycles with heterogeneous intermediaries.” *Review of Economic Studies* 91 (2): 817–857.

Corbae, Dean, and Pablo D’Erasmo. 2021. “Capital buffers in a quantitative model of banking industry dynamics.” *Econometrica* 89 (6): 2975–3023.

Corbae, Dean, and Ross Levine. 2018. “Competition, stability, and efficiency in financial markets.” In *Jackson Hole Symposium: Changing Market Structure and Implications for Monetary Policy, Kansas City Federal Reserve.*

Di Tella, Sebastian, and Pablo Kurlat. 2021. “Why are banks exposed to monetary policy?” *American Economic Journal: Macroeconomics* 13 (4): 295–340.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl. 2017. “The deposits channel of monetary policy.” *The Quarterly Journal of Economics* 132 (4): 1819–1876.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl. 2018. “A model of monetary policy and risk premia.” *The Journal of Finance* 73 (1): 317–373.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl. 2021. “Banking on deposits: Maturity transformation without interest rate risk.” *The Journal of Finance* 76 (3): 1091–1143.

Egan, Mark, Ali Hortaçsu, and Gregor Matvos. 2017. “Deposit competition and financial fragility: Evidence from the us banking sector.” *American Economic Review* 107 (1): 169–216.

Eggertsson, Gauti B, Ragnar E Juelsrud, Lawrence H Summers, and Ella Getz Wold. 2024. “Negative nominal interest rates and the bank lending channel.” *Review of Economic Studies* 91 (4): 2201–2275.

English, William B, Skander J Van den Heuvel, and Egon Zakrajšek. 2018. “Interest rate risk and bank equity valuations.” *Journal of Monetary Economics* 98 80–97.

Feenstra, Robert C. 1986. “Functional equivalence between liquidity costs and the utility of money.” *Journal of Monetary Economics* 17 (2): 271–291.

Gabaix, Xavier, and Ralph SJ Koijen. 2021. “In search of the origins of financial fluctuations: The inelastic markets hypothesis.”Technical report, National Bureau of Economic Research.

Gali, Jordi. 1999. “Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations?” *American economic review* 89 (1): 249–271.

Gerali, Andrea, Stefano Neri, Luca Sessa, and Federico M Signoretti. 2010. “Credit and Banking in a DSGE Model of the Euro Area.” *Journal of money, Credit and Banking* 42 107–141.

Gertler, Mark, and Peter Karadi. 2011. “A model of unconventional monetary policy.” *Journal of monetary Economics* 58 (1): 17–34.

Gertler, Mark, and Nobuhiro Kiyotaki. 2010. "Financial intermediation and credit policy in business cycle analysis." In *Handbook of monetary economics*, Volume 3. 547–599, Elsevier.

Gomez, Matthieu, Augustin Landier, David Sraer, and David Thesmar. 2021. "Banks' exposure to interest rate risk and the transmission of monetary policy." *Journal of Monetary Economics* 117 543–570. https://doi.org/10.1016/j.jmoneco.2020.03.011.

Gourinchas, Pierre-Olivier, Walker Ray, and Dimitri Vayanos. 2022. "A preferred-habitat model of term premia, exchange rates, and monetary policy spillovers."Technical report, National Bureau of Economic Research.

Hansen, Lars, and Thomas Sargent. 1981. "Exact linear rational expectations models: specification and estimation."Technical report, Federal Reserve Bank of Minneapolis.

He, Zhiguo, and Arvind Krishnamurthy. 2013. "Intermediary asset pricing." *American Economic Review* 103 (2): 732–770.

Herbst, Edward P, and Frank Schorfheide. 2016. *Bayesian estimation of DSGE models*. Princeton University Press.

Jamilov, Rustam, and Tommaso Monacelli. 2023. "Bewley banks." *Available at SSRN 3732172*.

Kashyap, Anil K, and Jeremy C Stein. 1995. "The impact of monetary policy on bank balance sheets." In *Carnegie-rochester conference series on public policy*, Volume 42. 151–195, Elsevier.

Kashyap, Anil K, and Jeremy C Stein. 2000. "What do a million observations on banks say about the transmission of monetary policy?" *American Economic Review* 90 (3): 407–428.

Kimball, Miles. 1995. "The Quantitative Analytics of the Basic Neomonetarist Model." *Journal of Money, Credit and Banking* 27 (4): 1241–77.

Kiyotaki, Nobuhiro, and John Moore. 1997. "Credit cycles." *Journal of political economy* 105 (2): 211–248.

Klein, Michael A. 1971. "A theory of the banking firm." *Journal of money, credit and banking* 3 (2): 205–218.

Koijen, Ralph SJ, and Motohiro Yogo. 2019. "A demand system approach to asset pricing." *Journal of Political Economy* 127 (4): 1475–1515.

Kurlat, Pablo. 2019. "Deposit spreads and the welfare cost of inflation." *Journal of Monetary Economics* 106 78–93.

Liu, Laura, and Mikkel Plagborg-Møller. 2023. "Full-information estimation of heterogeneous agent models using macro and micro data." *Quantitative Economics* 14 (1): 1–35.

Monti, Mario et al. 1972. *Deposit, credit and interest rate determination under alternative bank objective function*. North-Holland/American Elsevier Amsterdam.

Ogawa, Toshiaki. 2025. "Understanding The Great Recession Through The Banking Sector." *International Economic Review*.

Plagborg-Møller, Mikkel. 2019. "Bayesian inference on structural impulse response functions." *Quantitative Economics* 10 (1): 145–184.

Rotemberg, Julio J. 1982. "Sticky prices in the United States." *Journal of political economy* 90 (6): 1187–1211.

Scharfstein, David, and Adi Sunderam. 2016. "Market power in mortgage lending and the transmission of monetary policy." *Unpublished working paper. Harvard University* 2.

Smets, Frank, and Rafael Wouters. 2007. "Shocks and frictions in US business cycles: A Bayesian DSGE approach." *American economic review* 97 (3): 586–606.

Ulate, Mauricio. 2021. "Going Negative at the Zero Lower Bound: The Effects of Negative Nominal Interest Rates." *American Economic Review* 111 (1): 1–40.

Vayanos, Dimitri, and Jean-Luc Vila. 2021. "A preferred-habitat model of the term structure of interest rates." *Econometrica* 89 (1): 77–112.

Wang, Olivier. 2018. "Banks, low interest rates, and monetary policy transmission." *NYU Stern School of Business.*

Wang, Yifei, Toni M Whited, Yufeng Wu, and Kairong Xiao. 2022. "Bank market power and monetary policy transmission: Evidence from a structural estimation." *The Journal of Finance* 77 (4): 2093–2141.

Whited, Toni M, Yufeng Wu, and Kairong Xiao. 2021. "Low interest rates and risk incentives for banks with market power." *Journal of Monetary Economics* 121 155–174.

Appendix

A Appendix for Macro Equilibrium System

A.1 Optimality conditions, steady state, and log-linearization

A.1.1 Households

We obtain the following optimality conditions for the demand for risk-free bonds, deposit demand, consumption goods demand, and labor supply by defining the marginal utility of consumption as $\Lambda_t \equiv \frac{\partial U(C_t, C_{t-1}, H_t)}{\partial C_t}$:

$$\begin{aligned}
[B_t]: &\quad \beta \mathbb{E}_t \left[\frac{1+r_t^b}{1+\pi_{t+1}} \Lambda_{t+1} \right] = \Lambda_t \\
[D_t]: &\quad \beta \mathbb{E}_t \left[\frac{1+r_t^d}{1+\pi_{t+1}} \Lambda_{t+1} \right] - e^{v_t^d} \zeta \left(D_t - D^* \right) = \Lambda_t \\
[C_t]: &\quad \Lambda_t = e^{v_t^p} \frac{1}{C_t - hC_{t-1}} - \mathbb{E}_t \left[e^{v_{t+1}^p} \frac{\beta h}{C_{t+1} - hC_t} \right]
\end{aligned} \tag{A.1}$$

The liquidity services from demand generate a wedge between the deposit rate and the short-term rate:

$$\beta \mathbb{E}_t \left[\frac{r_t^b - r_t^d}{1+\pi_{t+1}} \Lambda_{t+1} \right] = -e^{v_t^d} \zeta \left(D_t - D^* \right). \tag{A.2}$$

We assume a zero-inflation steady state when solving the model. Thus we have:

$$\begin{aligned}
[B_t]: &\quad 1 = \beta \left(1 + r^b \right) \\
[D_t]: &\quad 1 = \frac{1+r^d}{1+r^b} - \frac{\zeta}{\Lambda} \left(D - D^* \right) \\
[C_t]: &\quad \Lambda = \frac{1-\beta h}{1-h} \frac{1}{C}
\end{aligned} \tag{A.3}$$

Log-linearizing around the steady state, we obtain:

$$\begin{aligned}
[B_t]: &\quad r_t^b - \mathbb{E}_t \pi_{t+1} + \mathbb{E}_t \lambda_{t+1} = \lambda_t \\
[D_t]: &\quad \beta \left(1 + r^d \right) \left[r_t^d - \mathbb{E}_t \pi_{t+1} + \mathbb{E}_t \lambda_{t+1} \right] - \frac{\zeta}{\Lambda} \left(D - D^* \right) v_t^d - \frac{\zeta}{\Lambda} D d_t = \lambda_t \\
[C_t]: &\quad \lambda_t = \frac{\beta h}{(1-\beta h)(1-h)} \mathbb{E}_t c_{t+1} - \frac{1+\beta h^2}{(1-\beta h)(1-h)} c_t + \frac{h}{(1-\beta h)(1-h)} c_{t-1} + \frac{1-\rho_p \beta h}{1-\beta h} v_t^p
\end{aligned} \tag{A.4}$$

A.1.2 Labor unions

We assume nominal rigidities in wage setting with Rotemberg (1982)-like disutility. We follow Auclert et al. (2024) for the derivation of the nonlinear wage Phillips curve.[24] Auclert et al. (2024) derives the nonlinear wage Phillips curve for a heterogeneous agent model. By replacing heterogenous agents with a representative agent, we derive the following nonlinear wage Phillips curve:

$$\pi_t^w\left(1+\pi_t^w\right)=\frac{\varepsilon_t^w}{\kappa^w}\left(H_t\frac{\partial U\left(C_t,C_{t-1},H_t\right)}{\partial H_t}-\frac{W_t^r H_t}{\mathcal{M}_t^w}\frac{\partial U\left(C_t,C_{t-1},H_t\right)}{\partial C_t}\right)+\beta\pi_{t+1}^w\left(1+\pi_{t+1}^w\right), \tag{A.5}$$

where $1+\pi_t^w\equiv\frac{W_t}{W_{t-1}}=\left(1+\pi_t\right)\frac{W_t^r}{W_{t-1}^r}$ and $\mathcal{M}_t^w=\frac{\varepsilon_t^w}{\varepsilon_t^w-1}$. Next, we obtain the following expression, found in the main text:

$$\pi_t^w\left(1+\pi_t^w\right)=\frac{\varepsilon_t^w}{\kappa^w}\left(e^{v_t^p}\chi H_t^{1+\varphi}-\frac{\Lambda_t W_t^r H_t}{\mathcal{M}_t^w}\right)+\beta\pi_{t+1}^w\left(1+\pi_{t+1}^w\right). \tag{A.6}$$

Log-linearizing around the zero-wage-inflation steady state, we obtain the following linear wage Phillips curve:

$$\pi_t^w=\frac{\varepsilon^w}{\kappa^w}\chi H^{1+\varphi}\left[\mu_t^w+v_t^p+\varphi h_t-\lambda_t-w_t^r\right]+\beta\mathbb{E}_t\pi_{t+1}^w. \tag{A.7}$$

A.1.3 Intermediate-goods-producing firms

The problem of the intermediate-goods-producing firms is:

$$\max_{K_t,H_t}\frac{P_t^{ig}}{P_t}A_t\left(K_t^\alpha H_t^{1-\alpha}\right)^\nu+(1-\delta)Q_tK_t-\frac{1+r_{t-1}^l}{1+\pi_t}Q_{t-1}K_t-\frac{W_t}{P_t}H_t. \tag{A.8}$$

The optimal demand for capital and labor are:

$$\begin{aligned}
[K_t]:&\quad \nu\alpha\frac{P_t^{ig}}{P_t}A_t\left(K_t^\alpha H_t^{1-\alpha}\right)^\nu=\left[\frac{1+r_{t-1}^l}{1+\pi_t}Q_{t-1}-\left(1-\delta\right)Q_t\right]K_t\\
[H_t]:&\quad \nu(1-\alpha)\frac{P_t^{ig}}{P_t}A_t\left(K_t^\alpha H_t^{1-\alpha}\right)^\nu=\frac{W_t}{P_t}H_t
\end{aligned} \tag{A.9}$$

Solving the above equation with respect to the loan rate, r_t^l, we obtain:

$$\frac{1+r_{t-1}^l}{1+\pi_t}Q_{t-1}=\nu\alpha\frac{P_t^{ig}}{P_t}A_t\frac{H_t^{(1-\alpha)\nu}}{K_t^{1-\alpha\nu}}+\left(1-\delta\right)Q_t. \tag{A.10}$$

If we solve the above equation with respect to H_t, we obtain:

$$H_t=\left[\frac{\nu(1-\alpha)P_t^{ig}A_t\left(K_t\right)^{\alpha\nu}}{W_t}\right]^{\frac{1}{1-(1-\alpha)\nu}}. \tag{A.11}$$

[24] See Auclert et al. (2024) and Auclert et al. (2021) for the details of the derivation of the nonlinear wage Phillips curve in a heterogenous agent model.

Note that the profit of the intermediate-goods-producing firm is:

$$\Pi_t^{ig} = (1-\nu)\,\frac{P_t^{ig}}{P_t} Y_t. \tag{A.12}$$

Therefore, the profit of intermediate-goods-producing firms vanishes once we assume constant returns to scale in production. The steady-state values for the loan rate and labor demand are given by:

$$r^l = \nu\alpha A \frac{\varepsilon-1}{\varepsilon}\frac{H^{(1-\alpha)\nu}}{K^{1-\alpha\nu}} - \delta, \quad H = \left[\nu(1-\alpha)A\frac{P^{ig}}{W}K^{\alpha\nu}\right]^{\frac{1}{1-(1-\alpha)\nu}}. \tag{A.13}$$

Log-linearizing around the steady state yields:

$$\left(1+r^l\right)\left(r_{t-1}^l - \pi_t\right) = \left(r^l+\delta\right)\left[a_t + p_t^{ig,r} - q_{t-1} + (1-\alpha)\nu h_t + (1-\alpha\nu)k_t\right] + (1-\delta)\left(q_t - q_{t-1}\right) \tag{A.14}$$

$$h_t = \frac{1}{1-(1-\alpha)\,\nu}\left[a_t + p_t^{ig,r} + \alpha\nu k_t - w_t^r\right] \tag{A.15}$$

A.1.4 Final good producer and retailers

A representative final good producer aggregates differentiated varieties supplied by monopolistic retailers at prices P_t^{ig} and produces the final good at a competitive price P_t. The monopolistic retailers face a conventional Rotemberg-type adjustment cost. The optimization problem for them is:

$$\max_{p_t(i)} \sum_{t=0}^{\infty} \beta^t \Lambda_t \left[\frac{p_t\,(i) - P_t^{ig}}{P_t} y_t\,(i) - \frac{\kappa^p}{2}\left(\frac{p_t\,(i)}{p_{t-1}\,(i)} - 1\right)^2 \frac{p_t\,(i)}{P_t} y_t\,(i)\right]. \tag{A.16}$$

Solving the above problem with respect to $p_t\,(i)$ results in:

$$\left[(1-\varepsilon_t^p) + \varepsilon_t^p \frac{P_t^{ig}}{P_t} - \kappa^p\left(\Pi_t - 1\right)\Pi_t\right] + \beta\kappa^p \frac{\Lambda_{t+1}}{\Lambda_t}\left(\Pi_{t+1}-1\right)\Pi_{t+1}\frac{Y_{t+1}}{Y_t} = 0 \tag{A.17}$$

Defining the markup of final good producers as $\mathcal{M}_t \equiv \frac{\varepsilon_t^p}{\varepsilon_t^p - 1}$, we obtain the following nonlinear NK Phillips curve:

$$\left[1 - \mathcal{M}_t P_t^{ig,r} - \frac{\kappa^p}{1-\varepsilon_t^p}\left(\Pi_t - 1\right)\Pi_t\right] + \frac{\beta\kappa^p}{1-\varepsilon_t^p}\frac{\Lambda_{t+1}}{\Lambda_t}\left(\Pi_{t+1}-1\right)\Pi_{t+1}\frac{Y_{t+1}}{Y_t} = 0. \tag{A.18}$$

We assume zero inflation at the steady state when deriving the NK Phillips curve. This results in the following real price of intermediate goods:

$$P^{ig,r} = \frac{1}{\mathcal{M}} = \frac{\varepsilon^p - 1}{\varepsilon^p}. \tag{A.19}$$

Log-linearizing around the steady state and rearranging the equation above yields the following expression for the linearized NK Phillips curve, as presented in the main text:

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \frac{\varepsilon^p - 1}{\kappa^p} p_t^{ig,r} + \frac{\varepsilon^p - 1}{\kappa^p} \mu_t. \tag{A.20}$$

A.1.5 Law of motion of capital

Capital is accumulated according to the following law of motion:

$$K_{t+1} = (1-\delta)K_t + I_t \left(1 - \Xi_i \left(\frac{I_t}{I_{t-1}}\right)\right). \tag{A.21}$$

At the steady state, investment equals depreciated capital, $\delta K = I$. This results in the following linear equation for capital accumulation:

$$k_{t+1} = (1-\delta)k_t + \delta i_t. \tag{A.22}$$

A.1.6 Capital good producers

The problem faced by capital producers is:

$$\max_{I_t} \sum_{t=0}^{\infty} \beta^t \Lambda_t \left(Q_t I_t \left(1 - \Xi_i \left(\frac{I_t}{I_{t-1}}\right)\right) - I_t \right). \tag{A.23}$$

Solving this equation yields:

$$\Lambda_t \left(Q_t \left[1 - \frac{\kappa^i}{2} \left(\frac{I_t}{I_{t-1}} - 1 \right)^2 - \kappa^i \frac{I_t}{I_{t-1}} \left(\frac{I_t}{I_{t-1}} - 1 \right) \right] - 1 \right) + \beta \Lambda_{t+1} Q_{t+1} \kappa^i \left(\frac{I_{t+1}}{I_t} \right)^2 \left(\frac{I_{t+1}}{I_t} - 1 \right) = 0. \tag{A.24}$$

At the steady state, the price of capital is $Q = 1$. Log-linearizing yields:

$$q_t = \kappa^i \left[-\beta i_{t+1} + (\beta + 1) i_t - i_{t-1} \right]. \tag{A.25}$$

A.1.7 Monetary policy

Taylor rule The monetary authority controls the short-term interest rate according to the conventional Taylor rule. We assume that the monetary authority responds to both inflation and the output gap.

$$\frac{1+r_t^b}{1+r^b} = \left(\frac{1+r_{t-1}^b}{1+r^b} \right)^{\rho_{mp}} \left[\left(\frac{1+\pi_t}{1+\pi} \right)^{\phi^\pi} \left(\frac{Y_t}{Y} \right)^{\phi^y} \right]^{(1-\rho_{mp})} \exp\left(\sigma_{mp} \epsilon_t^{mp}\right) \quad \text{where} \quad \epsilon_t^{mp} \sim N(0,1). \tag{A.26}$$

Around the zero-inflation steady state, we have the following linearized version of the Taylor rule:

$$r_t^b = \rho_{mp} r_{t-1}^b + \left(1 - \rho_{mp}\right) \left(\phi^\pi \pi_t + \phi^y y_t\right) + \sigma_{mp} \epsilon_t^{mp}. \tag{A.27}$$

Asset purchases by the central bank We model the impact of long-term bond purchases by the central bank on the long-term bond yield. We employ a market-segmentation assumption on the price determination of the long-term bond following Vayanos and Vila (2021). We assume that non-bank financial intermediaries (NBFI) and the central bank have the lion's share of long-term bonds. This assumption implies that commercial banks are price takers when purchasing long-term bonds.

The NBFIs are risk averse and maximize a mean-variance objective function. Our formulation is based on Campbell and Viceira (2002). Assuming that the investor trades off the mean and variance in a linear fashion, the investor's maximization problem is:

$$\max_{\boldsymbol{\alpha}_t} \boldsymbol{\alpha}_t' \left(\mathbb{E}_t \mathbf{R}_{t+1} - R_t^f \boldsymbol{\iota} \right) - \frac{\omega_t}{2} \boldsymbol{\alpha}_t' \boldsymbol{\Sigma}_t \boldsymbol{\alpha}_t, \tag{A.28}$$

where $\boldsymbol{\alpha}_t$ is a vector of allocations to the risky assets, $\boldsymbol{\iota}$ is a vector of ones, and $\boldsymbol{\Sigma}_t$ is a variance-covariance matrix of the return on the risky assets. $\mathbf{R}_t$ is the return on risky assets and R_t^f is the return on the risk-free bond. ω_t is the risk-aversion parameter. The solution to the maximization problem is:

$$\boldsymbol{\alpha}_t = \frac{1}{\omega_t} \boldsymbol{\Sigma}_t^{-1} \left(\mathbb{E}_t \mathbf{R}_{t+1} - R_t^f \boldsymbol{\iota} \right). \tag{A.29}$$

In our model, NBFIs' portfolios consist of short-term bonds and long-term bonds with a zero net position.

$$P_t^m H_t^{10y} = \frac{R_t^{10y} - R_t^b}{\omega e^{v_t^{mrt}} \Sigma}, \tag{A.30}$$

where the return on the 10-year bond is defined as:

$$R_t^{10y} = \frac{1 + \delta P_t^m}{P_{t-1}^m}. \tag{A.31}$$

δ controls the maturity of all subsets of long-term government bonds. $e^{v_t^{mrt}}$ controls the extent to which NBFIs take long-term risk. We assume that this long-term risk is captured by the following AR(1) process:

$$v_t^{mrt} = \rho_\xi v_{t-1}^{mrt} + \sigma_{v^{mrt}} \varepsilon_t^{v^{mrt}}, \quad \varepsilon_t^{v^{mrt}} \sim \text{iid}(0,1). \tag{A.32}$$

Since we assume that 10-year government bonds are supplied inelastically by the fiscal authority and that supply remains unchanged, the market-clearing condition for the long-term bond is:

$$G^{10y} = CB_t^{10y} + H_t^{10y}. \tag{A.33}$$

Log-linearizing eqn. (A.30), eqn. (A.31), and eqn. (A.33), and combining them yields:

$$\left(1 + \frac{r^{10y} - r^b}{1 + r^{10y} - \delta}\right) r_t^{10y} = \frac{1 + r^b}{1 + r^{10y}} r_t^b - \frac{r^{10y} - r^b}{1 + r^{10y}} \frac{CB^{10y}}{H^{10y}} cb_t^{10y} + \frac{r^{10y} - r^b}{1 + r^{10y}} v_t^{mrt}, \tag{A.34}$$

where terms without time subscripts denote steady-state values. Finally, we obtain the following equation for the determination of the price of the long-term bond, as presented in the main text:

$$r_t^{10y} = \iota^{st} \cdot r_t^b - \iota^{cb} \cdot cb_t^{10y} + v_t^{mrt}, \tag{A.35}$$

where ι^{st} stands for the elasticity of the long-term government bond yield with respect to the short-term government bond yield, and ι^{cb} stands for the elasticity of the long-term yield with respect to the central bank's asset purchases. cb_t^{10y} denotes a long-term bond purchase shock by the central bank. Thus, a positive value of cb_t^{10y} indicates a quantitative easing shock, which lowers the yield of the long-term bond. Once the NBFIs become risk-averse, which implies an increase in v_t^{mrt}, the long-term rate increases. As v_t^{mrt} is the exogenous shock in our model, we can disregard its coefficient.

Lastly, the profit earned by NBFIs is transfered to the households and amounts to:

$$\Pi_t^{FI} = \frac{1 + r_{t-1}^{10y}}{1 + \pi_t} H_{t-1}^{10y} + \frac{1 + r_{t-1}^{b}}{1 + \pi_t} H_{t-1}. \tag{A.36}$$

A.1.8 Fiscal policy

We assume that the government levies a lump-sum tax, T_t, to satisfy the following budget constraint:

$$\frac{1 + r_{t-1}^{b}}{1 + \pi_t} G_{t-1} + \frac{1 + r_{t-1}^{10y}}{1 + \pi_t} G^{10y} + T_t \leq G_t + G^{10y}. \tag{A.37}$$

The short-term bond is held by households, financial intermediaries, and the central bank. The long-term bond is mainly held by financial intermediaries and the central bank.

$$G_t = B_t + CB_t + H_t \quad \text{and} \quad G^{10y} = CB_t^{10y} + H_t^{10y}. \tag{A.38}$$

Note that banks also invest in long-term securities, but we assume that the relative size of bank holdings is infinitesimal.

A.1.9 Walras's law

In this subsection, we check if Walras's law holds in this economy when the central bank purchases assets. We start with the households' budget constraint:

$$\begin{aligned} C_t + B_t + D_t \leq \frac{W_t}{P_t} H_t + \frac{1 + r_{t-1}^{b}}{1 + \pi_t} B_{t-1} + \frac{1 + r_{t-1}^{d}}{1 + \pi_t} D_{t-1} \\ + \Pi_t^{ig} + \Pi_t^{F} + \Pi_t^{K} + \Pi_t^{B} + \Pi_t^{FI} + T_t. \end{aligned} \tag{A.39}$$

The government's budget constraint is:

$$\frac{1 + r_{t-1}^{b}}{1 + \pi_t} G_{t-1} + \frac{1 + r_{t-1}^{10y}}{1 + \pi_t} G^{10y} + T_t \leq G_t + G^{10y} \tag{A.40}$$

The central bank's budget constraint is:

$$CB_t + CB_t^{10y} \leq \frac{1 + r_{t-1}^{b}}{1 + \pi_t} CB_{t-1} + \frac{1 + r_{t-1}^{10y}}{1 + \pi_t} CB_{t-1}^{10y} + \Delta M_t \tag{A.41}$$

From the input to the production function, we have:

$$\frac{P_t^{ig}}{P_t}\nu\left(1-\alpha\right)Y_t = \frac{W_t}{P_t}H_t \quad \text{and}$$

$$\frac{P_t^{ig}}{P_t}\nu\alpha Y_t = \left[\frac{1+r_{t-1}^l}{1+\pi_t}Q_{t-1} - \left(1-\delta\right)Q_t\right]K_t = \frac{1+r_{t-1}^l}{1+\pi_t}L_{t-1} - \left[K_{t+1} - I_t\left(1-\Xi_i\left(\frac{I_t}{I_{t-1}}\right)\right)\right]Q_t \tag{A.42}$$

Profits obtained by the intermediate-goods-producing firms under the assumption of decreasing returns to scale are:

$$\Pi_t^{ig} = \left(1-\nu\right)\frac{P_t^{ig}}{P_t}Y_t \tag{A.43}$$

Profits from retailers and capital producers are:

$$\Pi_t^F = \frac{P_t - P_t^{ig}}{P_t}Y_t - \Xi_p\left(\frac{P_t}{P_{t-1}}\right)Y_t \quad \text{and} \quad \Pi_t^K = Q_t I_t\left(1-\Xi_i\left(\frac{I_t}{I_{t-1}}\right)\right) - I_t. \tag{A.44}$$

Aggregating the profits of the real economy, we have:

$$\begin{aligned}\frac{W_t}{P_t}H_t + \frac{1+r_{t-1}^l}{1+\pi_t}L_{t-1} - \left[K_{t+1} - I_t\left(1-\Xi_i\left(\frac{I_t}{I_{t-1}}\right)\right)\right]Q_t + \Pi_t^{ig} + \Pi_t^F + \Pi_t^K \\ = Y_t - \Xi_p\left(\frac{P_t}{P_{t-1}}\right)Y_t + Q_t I_t\left(1-\Xi_i\left(\frac{I_t}{I_{t-1}}\right)\right) - I_t.\end{aligned} \tag{A.45}$$

In the following, we denote the aggregate level of net worth, deposits, loans, and security investments of banks with capital letters. Then the sum of net worth and net interest income is:

$$N_t + NII_t = \frac{1+r_t^{bl}}{1+\pi_{t+1}}B_t^L + \frac{1+r_t^l}{1+\pi_{t+1}}L_t - \frac{1+r_t^d}{1+\pi_{t+1}}D_t \tag{A.46}$$

The law of motion of bank net worth is $N_{t+1} = \left(1-\gamma\right)\left(N_t + NII_t\right)$. The profits of banks are $\Pi_{t+1}^B = \gamma\left(N_t + NII_t\right)$. Combining these expressions results in the following:

$$\begin{aligned}B_{t+1}^L + L_{t+1} - D_{t+1} + \Pi_{t+1}^B = N_{t+1} + \Pi_{t+1}^B &= N_t + NII_t \\ &= \frac{1+r_t^{bl}}{1+\pi_{t+1}}B_t^L + \frac{1+r_t^l}{1+\pi_{t+1}}L_t - \frac{1+r_t^d}{1+\pi_{t+1}}D_t\end{aligned} \tag{A.47}$$

Lastly, we assume that the volume of security investment in the long-term bond by banks is infinitesimal. Aggregating the above equations for each sector yields:

$$C_t + I_t + \Xi_p\left(\frac{P_t}{P_{t-1}}\right)Y_t \leq Y_t \tag{A.48}$$

This is the resource constraint of the economy.

B Appendix for Banking System

In this section, we formulate markup and markdown in the banking system à la Atkeson and Burstein (2008).

B.1 Deposit market

B.1.1 Deposit demand with quality adjusted nested-CES demand function

The representative households demand deposit from each bank so as to minimize their expenditure on saving in deposit.

$$\min_{d_{ij,t}}\{D_t\} \quad \text{s.t.} \quad \int \sum_{i\in j} r^d_{ij,t} d_{ij,t} dj = S_t \tag{B.1}$$

where the aggregate structures are as follows:

$$D_t \equiv \left[\int d_{j,t}^{\frac{\theta^d+1}{\theta^d}} dj\right]^{\frac{\theta^d}{\theta^d+1}}, \quad d_{j,t} \equiv \left[\sum_{i\in j} \frac{1}{\xi^d_{ij}} d_{ij,t}^{\frac{\eta^d+1}{\eta^d}}\right]^{\frac{\eta^d}{\eta^d+1}} \tag{B.2}$$

Aggregate deposit is denoted in upper-case, and bank- and market-level in lower-case. The first order condition with respect to $d_{ij,t}$ implies:

$$\frac{\partial d_{jt}}{\partial d_{ijt}} \frac{\partial D_t}{\partial d_{jt}} = \lambda^d r^d_{ij,t}, \tag{B.3}$$

where λ^d is the Lagrange multiplier on the budget constraint of households. Rearranging the above equation yields:

$$\begin{aligned} & r^d_{ij,t} = \frac{1}{\lambda^d} \frac{\partial d_{jt}}{\partial d_{ijt}} \frac{\partial D_t}{\partial d_{jt}} \quad \Leftrightarrow \quad r^d_{ij,t} d_{ij,t} = \frac{1}{\lambda^d} \left(\frac{\partial d_{jt}}{\partial d_{ijt}} \frac{d_{ij,t}}{d_{j,t}}\right) \left(\frac{\partial D_t}{\partial d_{jt}} \frac{d_{j,t}}{D_t}\right) D_t \\ \Leftrightarrow \quad & \sum_{i\in j} r^d_{ij,t} d_{ij,t} = \frac{1}{\lambda^d} \left(\sum_{i\in j} \frac{\partial d_{jt}}{\partial d_{ijt}} \frac{d_{ij,t}}{d_{j,t}}\right) \left(\frac{\partial D_t}{\partial d_{jt}} \frac{d_{j,t}}{D_t}\right) D_t \quad \Leftrightarrow \quad r^d_{j,t} d_{j,t} = \frac{1}{\lambda^d} \left(\frac{\partial D_t}{\partial d_{jt}} \frac{d_{j,t}}{D_t}\right) D_t \\ \Leftrightarrow \quad & \int r^d_{j,t} d_{j,t} dj = \frac{1}{\lambda^d} \left(\int \frac{\partial D_t}{\partial d_{jt}} \frac{d_{j,t}}{D_t} dj\right) D_t \quad \Leftrightarrow \quad r^d_t D_t = \frac{1}{\lambda^d} D_t \quad \Leftrightarrow \quad r^d_t = \frac{1}{\lambda^d}, \end{aligned} \tag{B.4}$$

where we use the following expressions,

$$\begin{aligned} & D_t = \left[\int d_{j,t}^{\frac{\theta^d+1}{\theta^d}} dj\right]^{\frac{\theta^d}{\theta^d+1}} \quad \Leftrightarrow \quad \frac{\partial D_t}{\partial d_{j,t}} \frac{d_{j,t}}{D_t} = \left(\frac{d_{j,t}}{D_t}\right)^{\frac{\theta^d+1}{\theta^d}} \quad \Leftrightarrow \quad \int \frac{\partial D_t}{\partial d_{j,t}} \frac{d_{j,t}}{D_t} dj = 1, \\ & d_{j,t} = \left[\sum_{i\in j} \frac{1}{\xi^d_{ij}} d_{ij,t}^{\frac{\eta^d+1}{\eta^d}}\right]^{\frac{\eta^d}{\eta^d+1}} \quad \Leftrightarrow \quad \frac{\partial d_{j,t}}{\partial d_{ij,t}} \frac{d_{ij,t}}{d_{j,t}} = \frac{1}{\xi^d_{ij}} \left(\frac{d_{ij,t}}{d_{j,t}}\right)^{\frac{\eta^d+1}{\eta^d}} \quad \Leftrightarrow \quad \sum_{i\in j} \frac{\partial d_{j,t}}{\partial d_{ij,t}} \frac{d_{ij,t}}{d_{j,t}} = 1. \end{aligned} \tag{B.5}$$

Then, we obtain the following market-level deposit demand:

$$r^d_{j,t}d_{j,t} = \frac{1}{\lambda^d}\left(\frac{\partial D_t}{\partial d_{jt}}\frac{d_{j,t}}{D_t}\right)D_t \quad \Leftrightarrow \quad r^d_{j,t}d_{j,t} = \left(\frac{d_{j,t}}{D_t}\right)^{\frac{\theta^d+1}{\theta^d}} r^d_t D_t \quad \Leftrightarrow \quad d_{j,t} = \left(\frac{r^d_{j,t}}{r^d_t}\right)^{\theta^d} D_t. \quad \text{(B.6)}$$

In a similar way, we can also obtain bank-level deposit demand:

$$r^d_{ij,t}d_{ij,t} = \frac{1}{\lambda^d}\left(\frac{\partial d_{jt}}{\partial d_{ijt}}\frac{d_{ij,t}}{d_{j,t}}\right)\left(\frac{\partial D_t}{\partial d_{jt}}\frac{d_{j,t}}{D_t}\right)D_t \quad \Leftrightarrow \quad d_{ij,t} = \left(\xi^d_{ij}\frac{r^d_{ij,t}}{r^d_{j,t}}\right)^{\eta^d}\left(\frac{r^d_{j,t}}{r^d_t}\right)^{\theta^d} D_t. \quad \text{(B.7)}$$

Solving the equation for relative demand with respect to price yields the following demand function both at the market level and bank level:

$$\begin{aligned} r^d_{j,t} &= \left(\frac{d_{j,t}}{D_t}\right)^{\frac{1}{\theta^d}} r^d_t, \\ r^d_{ij,t} &= \frac{1}{\xi^d_{ij}}\left(\frac{d_{ij,t}}{d_{j,t}}\right)^{\frac{1}{\eta^d}} r^d_{j,t} = \frac{1}{\xi^d_{ij}}\left(\frac{d_{ij,t}}{d_{j,t}}\right)^{\frac{1}{\eta^d}}\left(\frac{d_{j,t}}{D_t}\right)^{\frac{1}{\theta^d}} r^d_t. \end{aligned} \quad \text{(B.8)}$$

B.1.2 Markdown in the deposit market

Since banks' deposit and lending market competitions are Cournot, banks choose deposit quantities taking their inverse deposit supply curve into account.[25] The optimality of each bank in deposit market can be abbreviated as follows:

$$d^*_{ij,t} = \arg\max_{d_{ij,t}}\left[mr^d_{ij,t}d_{ij,t} - r^d_{ij,t}\left(d_{ij,t}, d^*_{-ij,t}, r^d_t, D_t\right)d_{ij,t}\right] \quad \forall i \in j, \quad \text{(B.9)}$$

where

$$r^d_{ij,t}\left(d_{ij,t}, d^*_{-ij,t}, r^d_t, D_t\right) = \frac{1}{\xi^d_{ij}}\left(\frac{d_{ij,t}}{d_{j,t}}\right)^{\frac{1}{\eta^d}}\left(\frac{d_{j,t}}{D_t}\right)^{\frac{1}{\theta^d}} r^d_t, \quad d_{j,t} = \left[\frac{1}{\xi^d_{ij}}d_{ij,t}^{\frac{\eta^d+1}{\eta^d}} + \sum_{k\neq i}\frac{1}{\xi^d_{kj}}d^{*\,\frac{\eta^d+1}{\eta^d}}_{kj,t}\right]^{\frac{\eta^d}{\eta^d+1}}. \quad \text{(B.10)}$$

[25]Since we assume Cournot competition in deposit and lending markets, we only show the derivation under Cournot competition. If readers are interested in details of the case of Bertrand competition, see Amiti et al. (2019). Amiti et al. (2019) shows the qualitative similarity between the quantity (Cournot) and price (Bertrand) oligopolistic competition. Indeed, with the case of markup, they show that the expression for $\sigma_{ij,t}$ is a simple average of within and across substitution elasticity parameter in price competition rather than a harmonic average as in quantity competition and those have the same monotonicity properties.

$mr^d_{ij,t}$ is the marginal return from supplying deposit. Solving a maximization problem of banks, we obtain the following banks' optimality condition:

$$\begin{aligned} mr^d_{ij,t} &= \frac{\partial r^d_{ij,t}\left(d_{ij,t}, d^*_{-ij,t}, r^d_t, D_t\right)}{\partial d_{ij,t}} d_{ij,t} + r^d_{ij,t}\left(d_{ij,t}, d^*_{-ij,t}, r^d_t, D_t\right) \\ &= r^d_{ij,t}\left(d_{ij,t}, d^*_{-ij,t}, r^d_t, D_t\right)\left[\frac{\partial r^d_{ij,t}\left(d_{ij,t}, d^*_{-ij,t}, r^d_t, D_t\right)}{\partial d_{ij,t}} \frac{d_{ij,t}}{r^d_{ij,t}\left(d_{ij,t}, d^*_{-ij,t}, r^d_t, D_t\right)} + 1\right] \\ &= r^d_{ij,t}\left(d_{ij,t}, d^*_{-ij,t}, r^d_t, D_t\right)\left[\frac{\partial \log r^d_{ij,t}\left(d_{ij,t}, d^*_{-ij,t}, r^d_t, D_t\right)}{\partial \log d_{ij,t}} + 1\right]. \end{aligned} \tag{B.11}$$

It should be noted that $r^d_{ij,t}\left(d_{ij,t}, d^*_{-ij,t}, r^d_t, D_t\right)$ should be consistent with the following markdown expression and sales share in the market.

$$\begin{aligned} \frac{\partial \log r^d_{ij,t}\left(d_{ij,t}, d^*_{-ij,t}, r^d_t, D_t\right)}{\partial \log d_{ij,t}} &= \frac{1}{\eta^d} + \left(\frac{1}{\theta^d} - \frac{1}{\eta^d}\right) \frac{\partial \log d_{j,t}\left(d_{ij,t}, d^*_{-ij,t}\right)}{\partial \log d_{ij,t}} \\ &= \frac{1}{\eta^d} + \left(\frac{1}{\theta^d} - \frac{1}{\eta^d}\right) \frac{1}{\xi^d_{ij}} \left(\frac{d_{ij,t}}{d_{j,t}}\right)^{\frac{\eta^d+1}{\eta^d}} \\ &= \frac{1}{\eta^d} + \left(\frac{1}{\theta^d} - \frac{1}{\eta^d}\right) \boldsymbol{s^d_{ij,t}}, \end{aligned} \tag{B.12}$$

where we use the following computation and the formula for the sales share:

$$\begin{aligned} \frac{\partial \log d_{j,t}\left(d_{ij,t}, d^*_{-ij,t}\right)}{\partial \log d_{ij,t}} &= \frac{\partial \log d_{j,t}\left(d_{ij,t}, d^*_{-ij,t}\right)}{\partial d_{ij,t}} \frac{\partial d_{ij,t}}{\partial \log d_{ij,t}} = \frac{\eta_d}{1+\eta_d} \frac{\partial \log\left(\frac{1}{\xi^d_{ij}} d_{ij,t}^{\frac{\eta^d+1}{\eta^d}} + \sum_{i \neq i} \frac{1}{\xi^d_{kj}} d^{*\,\frac{\eta^d+1}{\eta^d}}_{kj,t}\right)}{\partial d_{ij,t}} d_{ij,t} \\ &= \frac{\eta_d}{1+\eta_d} \frac{1}{\xi^d_{ij}} \frac{1+\eta_d}{\eta_d} \frac{d_{ij,t}^{\frac{\eta^d+1}{\eta^d}-1}}{\frac{1}{\xi^d_{ij}} d_{ij,t}^{\frac{\eta^d+1}{\eta^d}} + \sum_{k \neq i} \frac{1}{\xi^d_{kj}} d^{*\,\frac{\eta^d+1}{\eta^d}}_{kj,t}} d_{ij,t} = \frac{1}{\xi^d_{ij}} \frac{d_{ij,t}^{\frac{\eta^d+1}{\eta^d}}}{d_{j,t}^{\frac{\eta^d+1}{\eta^d}}}, \end{aligned} \tag{B.13}$$

$$\boldsymbol{s^d_{ij,t}} \equiv \frac{r^d_{ij,t} d_{ij,t}}{\sum_{i \in j} r^d_{ij,t} d_{ij,t}} = \frac{\frac{1}{\xi^d_{ij}} \left(\frac{d_{ij,t}}{d_{j,t}}\right)^{\frac{1}{\eta^d}} \left(\frac{d_{j,t}}{D_t}\right)^{\frac{1}{\theta^d}} d_{ij,t}}{\sum_{i \in j} \frac{1}{\xi^d_{ij}} \left(\frac{d_{ij,t}}{d_{j,t}}\right)^{\frac{1}{\eta^d}} \left(\frac{d_{j,t}}{D_t}\right)^{\frac{1}{\theta^d}} d_{ij,t}} = \frac{\frac{1}{\xi^d_{ij}} d_{ij,t}^{\frac{1+\eta^d}{\eta^d}}}{\sum_{i \in j} \frac{1}{\xi^d_{ij}} d_{ij,t}^{\frac{1+\eta^d}{\eta^d}}} = \frac{1}{\xi^d_{ij}} \left(\frac{d_{ij,t}}{d_{j,t}}\right)^{\frac{1+\eta^d}{\eta^d}}. \tag{B.14}$$

Note that the sales share can be expressed in terms of individual deposit rate:

$$\boldsymbol{s^d_{ij,t}} = \frac{r^d_{ij,t} d_{ij,t}}{\sum_{i\in j} r^d_{ij,t} d_{ij,t}} = \frac{r^d_{ij,t}\left(\xi^d_{ij}\frac{r^d_{ij,t}}{r^d_{j,t}}\right)^{\eta^d}\left(\frac{r^d_{j,t}}{r^d_t}\right)^{\theta^d} D_t}{\sum_{i\in j} r^d_{ij,t}\left(\xi^d_{ij}\frac{r^d_{ij,t}}{r^d_{j,t}}\right)^{\eta^d}\left(\frac{r^d_{j,t}}{r^d_t}\right)^{\theta^d} D_t} = \frac{\xi^{d\,\eta^d}_{ij}\left(r^d_{ij,t}\right)^{1+\eta^d}}{\sum_{i\in j}\xi^{d\,\eta^d}_{ij}\left(r^d_{ij,t}\right)^{1+\eta^d}} = \xi^{d\,\eta^d}_{ij}\left(\frac{r^d_{ij,t}}{r^d_{j,t}}\right)^{1+\eta^d}. \tag{B.15}$$

Define the equilibrium inverse deposit supply elasticity $\sigma^{d*}_{ij,t}$ as

$$\sigma^{d*}_{ij,t} = \left[\frac{\partial \log r^d_{ij,t}\left(d_{ij,t}, d^*_{-ij,t}, r^d_t, D_t\right)}{\partial \log d_{ij,t}}\right]^{-1} = \left[\boldsymbol{s^d_{ij,t}}\frac{1}{\theta^d} + \left(1-\boldsymbol{s^d_{ij,t}}\right)\frac{1}{\eta^d}\right]^{-1}. \tag{B.16}$$

Defining the markdown ratio of each bank in deposit market as $\mu^{d*}_{ij,t} \equiv \frac{r^{d*}_{ij,t}}{mr^d_{ij,t}}$, we can get the following expression for markdown:

$$\mu^{d*}_{ij,t} = \frac{\sigma^{d*}_{ij,t}}{\sigma^{d*}_{ij,t}+1} = \frac{1}{1+\sigma^{d*\,-1}_{ij,t}} = \frac{1}{1+\left[\boldsymbol{s^d_{i,t}}\frac{1}{\theta^d} + \left(1-\boldsymbol{s^d_{i,t}}\right)\frac{1}{\eta^d}\right]}. \tag{B.17}$$

Then, we summarize the market level aggregates and economy level aggregats for price and quantity as the followings. The market level aggregates are:

$$r^d_{j,t} = \left[\sum_{i\in j}\left(\xi^d_{ij}\right)^{\eta^d}\left(r^d_{ij,t}\right)^{1+\eta^d}\right]^{\frac{1}{1+\eta^d}}, \quad d_{ij,t} = \left(\xi^d_{ij}\frac{r^d_{ij,t}}{r^d_{j,t}}\right)^{\eta^d} d_{j,t}. \tag{B.18}$$

The economy level aggregates are:

$$r^d_t = \left[\int r^{d\;1+\theta^d}_{j,t}\, dj\right]^{\frac{1}{1+\theta^d}}, \quad d_{j,t} = \left(\frac{r^d_{j,t}}{r^d_t}\right)^{\theta^d} D_t. \tag{B.19}$$

Finally, we assume that the markets are symmetric across regions, j, in the economy. In other words, we assume that $r^d_{j,t} = r^d_t$ and $d_{j,t} = d_t$. Then, we obtain the following expressions for optimal price and share for banks in the deposit market in the main text.

$$
\begin{aligned}
\text{deposit rate}: \quad & r^d_{i,t} = \mu^d_{i,t} mr^d_{i,t} \\
\text{markdown}: \quad & \mu^d_{i,t} = \frac{1}{1+\left[\boldsymbol{s^d_{i,t}}\frac{1}{\theta^d} + \left(1-\boldsymbol{s^d_{i,t}}\right)\frac{1}{\eta^d}\right]} \\
\text{deposit sales share}: \quad & \boldsymbol{s^d_{i,t}} = \xi_i^{d\eta^d}\left(\frac{r^d_{i,t}}{r^d_t}\right)^{1+\eta^d} \\
\text{deposit share}: \quad & s^d_{i,t} \equiv \frac{d_{i,t}}{D_t} = \xi_i^{d\eta^d}\left(\frac{r^d_{i,t}}{r^d_t}\right)^{\eta^d}
\end{aligned}
\tag{B.20}
$$

B.2 Loan market

B.2.1 Loan demand with quality adjusted nested-CES demand function

The optimality conditions for loan demand is similar to that of deposit demand. The representative intermediate input goos producing firms demand loan from each bank so as to maximize their borrowing given the market lending rate. on saving in deposit.

$$
\max_{l_{ij,t}} \{L_t\} \quad \text{s.t.} \quad \int \sum_{i\in j} r^l_{ijt} l_{ijt} dj = E_t \tag{B.21}
$$

where the aggregator takes a form of followings:

$$
L_t = \left[\int l_{j,t}^{\frac{\theta^l-1}{\theta^l}} dj\right]^{\frac{\theta^l}{\theta^l-1}}, \quad l_{j,t} = \left[\sum_{i\in j} \xi^l_{ij} l_{ij,t}^{\frac{\eta^l-1}{\eta^l}}\right]^{\frac{\eta^l}{\eta^l-1}} \tag{B.22}
$$

The first order condition with respect to $l_{ij,t}$ implies:

$$
\frac{\partial l_{jt}}{\partial l_{ijt}}\frac{\partial L_t}{\partial l_{jt}} = -\lambda^l r^l_{ij,t}, \tag{B.23}
$$

where λ^l is the Lagrange multiplier on the budget constraint of the intermediate goods producing firms. Rearranging the above equation yields:

$$
r^l_{ij,t} = -\frac{1}{\lambda^l}\frac{\partial l_{jt}}{\partial l_{ijt}}\frac{\partial L_t}{\partial l_{jt}} \quad \Leftrightarrow \quad r^l_t = -\frac{1}{\lambda^l}. \tag{B.24}
$$

Then, we obtain the following market-level loan demand:

$$
r^l_{j,t} l_{j,t} = -\frac{1}{\lambda^l}\left(\frac{\partial L_t}{\partial l_{jt}}\frac{l_{j,t}}{L_t}\right) L_t \quad \Leftrightarrow \quad l_{j,t} = \left(\frac{r^l_{j,t}}{R^l_t}\right)^{-\theta^l} L_t. \tag{B.25}
$$

In a similar way, we can also obtain bank-level loan demand:

$$r^l_{ij,t} l_{ij,t} = -\frac{1}{\lambda^l}\left(\frac{\partial l_{jt}}{\partial l_{ijt}}\frac{l_{ij,t}}{l_{j,t}}\right)\left(\frac{\partial L_t}{\partial l_{jt}}\frac{l_{j,t}}{L_t}\right) L_t \quad \Leftrightarrow \quad l_{ij,t} = \left(\frac{1}{\xi^l_{ij}}\frac{r^l_{ij,t}}{r^l_{j,t}}\right)^{-\eta^l}\left(\frac{r^l_{j,t}}{r^l_t}\right)^{-\theta^l} L_t. \tag{B.26}$$

Solving the equation for relative demand with respect to price yields the following demand function both at the market level and bank level:

$$\begin{aligned} r^l_{j,t} &= \left(\frac{l_{j,t}}{L_t}\right)^{-\frac{1}{\theta^l}} r^l_t, \\ r^l_{ij,t} &= \xi^l_{ij}\left(\frac{l_{ij,t}}{l_{j,t}}\right)^{-\frac{1}{\eta^l}} r^l_{j,t} = \xi^l_{ij}\left(\frac{l_{ij,t}}{l_{j,t}}\right)^{-\frac{1}{\eta^l}}\left(\frac{l_{j,t}}{L_t}\right)^{-\frac{1}{\theta^l}} r^l_t. \end{aligned} \tag{B.27}$$

B.2.2 Markup in the loan market

The derivation of markup in the loan market is analogous to that of in deposit market, so that we skip the detailed computational step in this section and just focus on the results of computation. The optimality problem of each bank in the lending market is:

$$l^*_{ij,t} = \arg\max_{l_{ij,t}}\left[r^l_t\left(l_{ij,t}, l^*_{-ij,t}, r^l_t, L_t\right) l_{ij,t} - mc^l_{ij,t} l_{ij,t}\right] \quad \forall i \in j, \tag{B.28}$$

where

$$r^l_t\left(l_{ij,t}, l^*_{-ij,t}, r^l_t, L_t\right) = \xi^l_{ij}\left(\frac{l_{ij,t}}{l_{j,t}}\right)^{-\frac{1}{\eta^l}}\left(\frac{l_{j,t}}{L_t}\right)^{-\frac{1}{\theta^l}} r^l_t, \quad l_{j,t} = \left[l_{ij,t}^{\frac{\eta^l-1}{\eta^l}} + \sum_{k\neq i} l^{*\,\frac{\eta^l-1}{\eta^l}}_{kj,t}\right]^{\frac{\eta^l}{\eta^l-1}}. \tag{B.29}$$

$mc^l_{ij,t}$ the marginal cost of supplying loan. Solving the profit maximization problem of banks, we obtain the following banks' optimality condition:

$$mc^l_{ij,t} = r^l_t\left(l_{ij,t}, l^*_{-ij,t}, r^l_t, L_t\right)\left[\frac{\partial \log r^l_t\left(l_{ij,t}, l^*_{-ij,t}, r^l_t, L_t\right)}{\partial \log l_{ij,t}} + 1\right], \tag{B.30}$$

$$\text{and} \quad \frac{\partial \log r^l_t\left(l_{ij,t}, l^*_{-ij,t}, r^l_t, L_t\right)}{\partial \log l_{ij,t}} = -\frac{1}{\eta^l} + \left(-\frac{1}{\theta^l} + \frac{1}{\eta^l}\right)\boldsymbol{s^l_{ij,t}}, \tag{B.31}$$

where the market sales share is:

$$\boldsymbol{s^l_{ij,t}} = \xi^l_{ij}\left(\frac{l_{ij,t}}{l_{j,t}}\right)^{-\frac{1-\eta^l}{\eta^l}} = \xi^l_{ij}{}^{\eta_l}\left(\frac{r^l_{ij,t}}{r^l_{j,t}}\right)^{1-\eta^l}. \tag{B.32}$$

Define the equilibrium inverse loan supply elasticity $\sigma^{l*}_{ij,t}$ as

$$\sigma^{l*}_{ij,t} = -\left[\frac{\partial \log r^l_t\left(l_{ij,t}, l^*_{-ij,t}, r^l_t, L_t\right)}{\partial \log l_{ij,t}}\right]^{-1} = \left[\boldsymbol{s^l_{ij,t}}\frac{1}{\theta^l} + \left(1 - \boldsymbol{s^l_{ij,t}}\right)\frac{1}{\eta^l}\right]^{-1}. \tag{B.33}$$

Defining the markup ratio of each bank in lending market as $\mu^{l*}_{ij,t} \equiv \frac{r^{l*}_{ij,t}}{mc^l_{ij,t}}$, then we have:

$$\mu^{l*}_{ij,t} = \frac{\sigma^{l*}_{ij,t}}{\sigma^{l*}_{ij,t} - 1} = \frac{1}{1 - {\sigma^{l*}_{ij,t}}^{-1}} = \frac{1}{1 - \left[\boldsymbol{s^l_{ij,t}}\frac{1}{\theta^l} + \left(1 - \boldsymbol{s^l_{ij,t}}\right)\frac{1}{\eta^l}\right]}. \tag{B.34}$$

Then, we obtain the expressions for market level aggregates of lending rate and loans outstanding:

$$r^l_{j,t} = \left[\sum_{i\in j}\left(\xi^l_{ij}\right)^{\eta^l} {r^l_{ij,t}}^{1-\eta^l}\right]^{\frac{1}{1-\eta^l}}, \quad l_{ij,t} = \left(\frac{1}{\xi^l_{ij}}\frac{r^l_{ij,t}}{r^l_{j,t}}\right)^{-\eta^l} l_{j,t}. \tag{B.35}$$

Subsequently, we also obtain the expression for economy aggregates of price and quantity as:

$$r^l_t = \left[\int {r^l_{j,t}}^{1-\theta^l} dj\right]^{\frac{1}{1-\theta^l}}, \quad l_{j,t} = \left(\frac{r^l_{j,t}}{r^l_t}\right)^{-\theta^l} L_t. \tag{B.36}$$

Finally, we assume that the markets are symmetric across regions, j, in the economy. Mathematically speaking, we assume that $r^l_{j,t} = r^l_t$ and $l_{j,t} = l_t$. Then, we obtain the following expressions for optimal price and share for banks in the loan market in the main text.

$$\begin{aligned}
\text{lending rate}: &\quad r^l_{i,t} = \mu^l_{i,t} mc^l_{i,t} \\
\text{markup}: &\quad \mu^l_{i,t} = \frac{1}{1 - \left[s^l_{i,t}\frac{1}{\theta^l} + \left(1 - s^l_{i,t}\right)\frac{1}{\eta^l}\right]} \\
\text{loan sales share}: &\quad \boldsymbol{s^l_{i,t}} = \xi^{l\,\eta_l}_i \left(\frac{r^l_{i,t}}{r^l_t}\right)^{1-\eta^l} \\
\text{loan share}: &\quad s^l_{i,t} = \xi^{l\,\eta_l}_i \left(\frac{r^l_{i,t}}{r^l_t}\right)^{-\eta^l}
\end{aligned} \tag{B.37}$$

B.3 Alternative specification for banking system

In the main text, we consider the marginal net non-interest expense of loans and the marginal net non-interest income of deposits as additional exogeneous elements for the marginal cost of lending and marginal revenue of deposits. However, we can obtain a similar expression by expoiting the traditional leverage constraint and liquidity constraint as follows.

Stage 1: competition in deposit and loan markets We formulate the banks' problem at each time period as follows:

$$\begin{aligned} nii_{i,t} = & \max_{d_{i,t}, l_{i,t}, \psi_{i,t}} \left[r^l_{i,t} l_{i,t} - r^{bl}_{i,t} l_{i,t} + r^b_t d_{i,t} - r^d_{i,t} d_{i,t} + r^b_t n_{i,t} \right], \\ \text{s.t.} \quad & b^l_{i,t} + l_{i,t} \le n_{i,t} + d_{i,t}, \quad b^l_{i,t} \ge \omega^d d_{i,t}, \quad \omega^l n_{i,t} \ge l_{i,t}, \end{aligned} \tag{B.38}$$

Here we consider two additional conditions compared to the main text: a liquidity constraint and a leverage constraint. Similar to the main text, we assume that banks are bound by their budget constraint. The level of securities investment in long-term bonds is determined by construction after the choice of deposits and outstanding loans. When financial frictions take this form, banks' lending cannot exceed a multiple of ω^l and their net worth. Such constraints, which resemble capital requirements implemented in practice, are typical in models where banks face moral hazard problems (e.g. Kiyotaki and Moore (1997) and Gertler and Karadi (2011)). In addition, a bank's holdings of liquid bonds should exceed a fraction ω^d of its deposit supply. These types of constraints are typically present in models where banks face random deposit outflows (e.g. Drechsler et al. (2018) and Bianchi and Bigio (2022)), but may also reflect regulatory requirements. Putting Lagrange multipliers on both liquidity and leverage constraints, we formulate the following problem:

$$\begin{aligned} & \arg\max_{d_{i,t}, l_{i,t}, \psi_{i,t}} \left[r^l_{i,t} l_{i,t} - r^{bl}_{i,t} l_{i,t} + r^b_t d_{i,t} - r^d_{i,t} d_{i,t} + r^b_t n_{i,t} + \kappa^d_i \left(b^l_{i,t} - \omega^d d_{i,t} \right) + \kappa^l_i \left(\omega^l n_{i,t} - l_{i,t} \right) \right]. \\ & \arg\max_{d_{i,t}, l_{i,t}, \psi_{i,t}} \left[r^l_{i,t} l_{i,t} - r^{bl}_{i,t} l_{i,t} + r^b_t d_{i,t} - r^d_{i,t} d_{i,t} + r^b_t n_{i,t} + \kappa^d_i d_{i,t} - \kappa^l_i l_{i,t} \right]. \\ & \arg\max_{d_{i,t}, l_{i,t}, \psi_{i,t}} \left[r^l_{i,t} l_{i,t} - \left(r^{bl}_{i,t} + \kappa^l_i \right) l_{i,t} + \left(r^b_t + \kappa^d_i \right) d_{i,t} - r^d_{i,t} d_{i,t} + r^b_t n_{i,t} \right]. \end{aligned} \tag{B.39}$$

We should note that the shadow prices for both liquidity and leverage constraints differ across banks. Rewriting the objective function yields:

$$\begin{aligned} & r^{bl}_{i,t} \left(n_{i,t} + d_{i,t} - l_{i,t} \right) + r^l_{i,t} l_{i,t} - r^d_{i,t} d_{i,t} + \kappa^d_i \left(n_{i,t} + d_{i,t} - l_{i,t} - \omega^d d_{i,t} \right) + \kappa^l_i \left(\omega^l n_{i,t} - l_{i,t} \right) \\ = & \underbrace{\left[r^b_t d_{i,t} - r^d_{i,t} d_{i,t} + \left(1 - \omega_d \right) \kappa^d_i d_{i,t} \right]}_{\text{deposit market}} + \underbrace{\left[r^l_{i,t} l_{i,t} - r^{bl}_{i,t} l_{i,t} - \left(\kappa^l_i + \kappa^d_i \right) l_{i,t} \right]}_{\text{lending market}} \\ & + \left(r^{bl}_{i,t} - r^b_t \right) d_{i,t} + \left(r^{bl}_{i,t} + \kappa^d_i + \kappa^l_i \omega^l \right) n_{i,t}. \end{aligned} \tag{B.40}$$

We assume that the duration of the base rate for the deposit market is the three-month short-term rate. Moreover, we assume that banks earn additional profits by taking duration risk. We can then separate the problems in the deposit market and in the lending market as the follows:

$$\text{deposit market:} \quad d^*_{i,t} = \arg\max_{d_{i,t}} \left[r^b_t d_{i,t} - r^d_{i,t} \left(d_{i,t}, d^*_{-i,t}, r^d_t, D_t \right) d_{i,t} + \left(1 - \omega_d \right) \kappa^d_i d_{i,t} \right], \tag{B.41}$$

$$\text{lending market:} \quad l^*_{i,t} = \arg\max_{l_{i,t}} \left[r^l_{i,t} \left(l_{i,t}, l^*_{-i,t}, r^l_t, L_t \right) l_{i,t} - r^{bl}_{i,t} l_{i,t} - \left(\kappa^l_i + \kappa^d_i \right) l_{i,t} \right], \tag{B.42}$$

where $*$ denotes the optimal deposit and loan supplies. Given the relative demand indicated in (4) and (12), each bank solves the optimization problems for the supply of deposits and the supply of loans.

Solving these problems yields similar expressions for the deposit rate and the loan rate, but differ in the marginal revenue of deposits and the marginal cost of loans as follows:

$$\begin{aligned} mr_{i,t}^{d} &= r_{t}^{b} + (1-\omega_{d})\,\kappa_{i}^{d}, \\ mc_{i,t}^{l} &= r_{i,t}^{bl} + \left(\kappa_{i}^{l} + \kappa_{i}^{d}\right). \end{aligned} \tag{B.43}$$

Thus, we can assume that the estimated $mnnii_{i,t}^{d}$ and $mnnie_{i,t}^{l}$ represent the wedges that arise from the leverage and liquidity constraints.

B.4 Details of micro calibration

B.4.1 Constructing the measure of deposit markdowns and credit markups

We basically follow the approach of Bellifemine et al. (2022) and Jamilov and Monacelli (2023) when constructing our measure of deposit markdowns and loan markups for the U.S. banking system.[26] However, we depart from Jamilov and Monacelli (2023) in the definition of deposit markdowns to be consistent with the definition in our HBANK model. As in the main text, the markdown on deposits in our model is defined by the following formula:

$$\mu_{i,t}^{d} = \frac{r_{i,t}^{d}}{mr_{i,t}^{d}} = \frac{r_{i,t}^{d}}{r_{t}^{b} + mnnii_{i,t}}. \tag{B.44}$$

The marginal return on deposits, which is the income that banks earn by holding deposits, is the sum of the return on safe assets and the marginal net non-interest income. The return on safe assets is assumed to be equal to the short-term interest rate. Marginal net non-interest income ($mnnii$) is defined as the difference between marginal non-interest expenses and marginal non-interest income. As discussed in the main text, each bank must maintain a certain amount of liquidity (government bonds) to be prepared for a deposit outflow. In addition to the short-term interest rate, the bank must earn the marginal return on a unit of deposits plus the liquidity premium resulting from this liquidity constraint. In our model, we assume that this liquidity premium is equal to the fees earned on the bank's deposit business. This assumption is motivated by the model setting in Abadi, Brunnermeier, and Koby (2023).[27]

Our methodology for constructing the measure of credit loan markups follows Bellifemine et al. (2022) and Jamilov and Monacelli (2023). In both deposit and loan markets, marginal non-interest expenses and marginal non-interest income are estimated for both deposit and loan markets using a log panel fixed effects regression. We rely on Jamilov and Monacelli (2023) for the exact estimation procedure and panel regression specification, but we summarize the datasets we use for our estimation in Table B.1.

[26] Jamilov and Monacelli (2023) follows Corbae and D'Erasmo (2021) in estimating banks' markups. However, Corbae and D'Erasmo (2021) does not estimate deposit markdowns. Thus, we follow Bellifemine et al. (2022) in estimating both deposit markdowns and loan markups.

[27] In Jamilov and Monacelli (2023), the markdown on deposits is defined as the ratio of the "safe return" (p) collected by the bank to the marginal cost (c) that the bank must bear to raise an additional unit of deposits and maintain its franchise. See Jamilov and Monacelli (2023) for more details.

Table B.1: Variable detail and source for micro calibration

Variable	Detail	Source
Assets	Total assets	Call Reports (RCFD2170)
Loans	Total loans and leases	Call Reports (RCFD2122-RCFD3123)
Deposits	Total domestic deposits	Call Reports (RCON2200)
Interest income on loans and leases	Total interest income on loans and leases	Call Reports (RIAD4010+RIAD4065)
Interest expense	Bank interest expenses on domestic deposits	Call Reports
	for 83Q1-09Q4	(RIAD4170-RIAD4172)
	for 10Q1-16Q4	(RIAD4172+RIADA517 +RIADA518+RIAD0093 +RIAD4508-RIAD4172)
	for 17Q1-	(RIAD4172+RIADHK04 +RIADHK03+RIAD0093 +RIAD4508-RIAD4172)
Non-interest expense	Bank non-interest expenses	Call Reports (RIAD4093)
Staff cost	Bank staff expenses	Call Reports (RIAD4135)
Securities	Bank holdings of securities	Call Reports
	for 76Q1-83Q4	(RCFD0400+RCFD0600 +RCFD0900+RCFD0950)
	for 84Q1-93Q4	(RCFD0390)
	for 94Q1-	(RCFD1754+RCFD1773)
Non-interest income	Bank non-interest income	Call Reports (RIAD4079)
Federal Fund rate	O/N	Federal Reserve Board
Market Yield on U.S. Treasury Securities	3-Month	FRED(DGS3MO)
	10-Year	FRED(DGS10)
Asset duration, ψ	Loans and leases with repricing maturity of	Call Reports
	less than three months	(RCFDA570+RCFDA564)
	more than three months and less than a year	(RCFDA571+RCFDA565)
	more than one year and less than three years	(RCFDA572+RCFDA566)
	more than three years and less than five years	(RCFDA573+RCFDA567)
	more than five years and less than fifteen years	(RCFDA574+RCFDA568)
	more than fifteen years	(RCFDA575+RCFDA569)

Figure B.1 shows the distribution of the estimated $mnnne^l_{2020}$ and the estimated $mnnni^d_{2020}$. We can see that these two values tend to be smaller for larger banks than smaller banks.

B.4.2 Loan duration

The duration risk of loans is calculated by weighting the maturity of each bank's corporate loans. Banks report their holdings of five asset categories in Call Reports (residential mortgage loans, all other loans, Treasuries and agency debt, MBS secured by residential mortgages, and other MBS). Since our interest is in corporate loans, we use *Loans and leases with repricing maturity of less than*

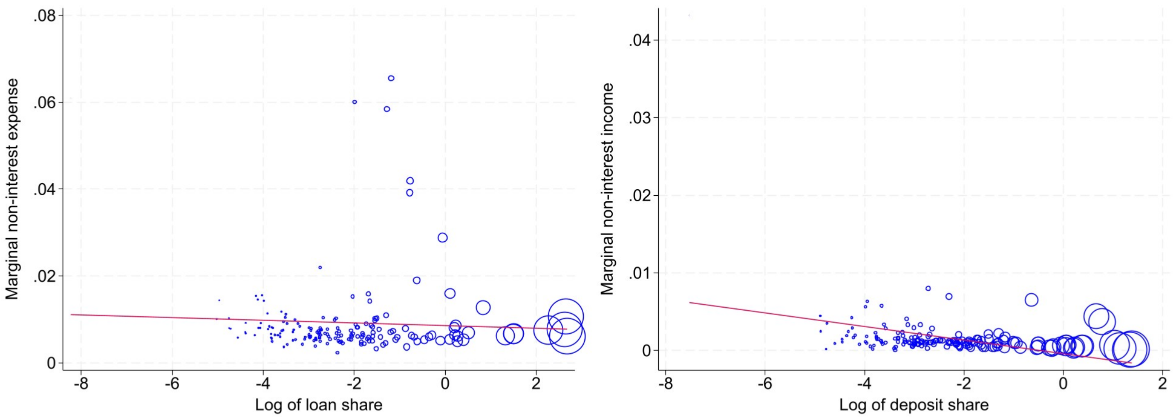

Figure B.1: The distribution of mnnie in loan market and mnnii in deposit market
NOTE: The horizontal axis shows the share of each bank and the size of each circle also represents the share of each bank.

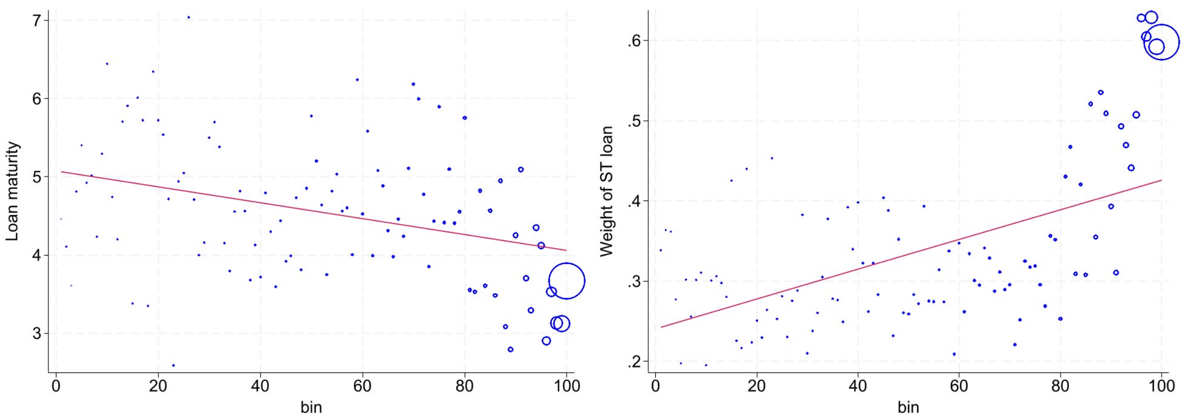

Figure B.2: Left: Loan repricing maturity, Right: Weight of short term loans
NOTE: We divide top 10% of U.S. banks into 100 bins and compare the weighted average of loan duration and share of short-term loans (i.e., less than 1 year).

three months. Each asset is broken down into six bins by repricing maturity interval (0 to 3 months, 3 to 12 months, 1 to 3 years, 3 to 5 years, 5 to 15 years, and over 15 years). To calculate the overall repricing maturity of a given asset category, we assign the interval midpoint to each bin (and 15 years to the last bin) and take a weighted average using the dollar amounts in each bin as weights.

In Figure B.2, we plot the duration of corporate loans and the fraction of shorter maturity loans (less than 1 year) for 100 bins. These 100 bins include non-community banks that have been in the top 10% of banks in terms of total assets at least once during the sample period. We sort these 600 banks, which account for 80% of total assets held by all banks, according to their loans outstanding and divide them into 100 bins (i.e., each bin includes 6 banks).

On the left panel in Figure B.2, we show that large banks tend to have shorter duration in their corporate loan portfolios. On the right panel in Figure B.2, we show that the fraction of shorter maturity (less than 1 year) loans is larger for large banks.

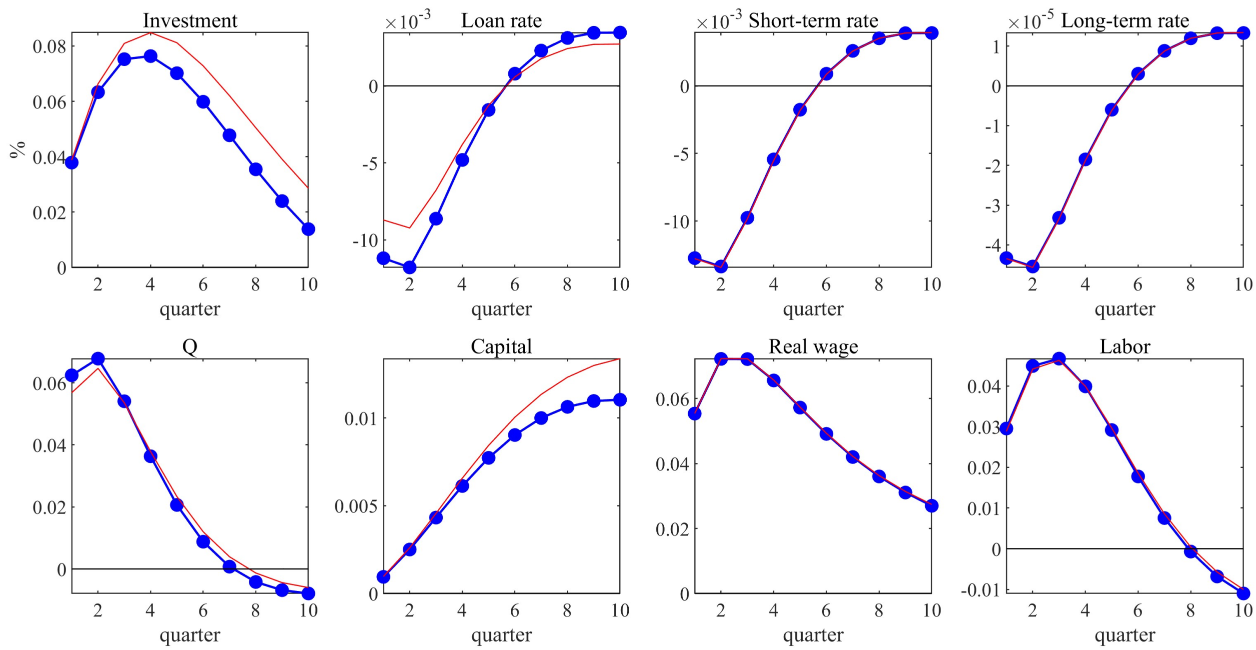

Figure C.1: IRFs to 1% negative markup shock

Note: We show the IRFs for the case when the AR (1) parameter of the price markup shock is $\rho_\mu = 0.5$ and the MA (1) parameter is $\theta_\mu = 0.0$. Blue solid circle lines represent the IRFs of the HBANK model and red solid lines represent the IRFs of the RBANK model.

C Appendix for Macro Analysis

C.1 Additional results for IRFs

Figure C.1 shows the IRFs in response to a negative price-markup shock.

C.2 Estimation

C.2.1 Data

Table C.1 summarizes the datasets used in our macro analysis.

The data used for the estimation include nine series: GDP, consumption, the inflation rate, the nominal wage inflation rate, the FF rate, the 10-year U.S.Treasury bond interest rate, the Fed's holdings of U.S.Treasury bonds, the lending rate, and deposits outstanding (y, c, π, π^w, r_t^b, r_t^{10y}, cb, r_t^l, d). The data period is the same as for the micro-moment analysis - 35 years (140 quarters), from 1985 to 2020.

As described in the main text, for GDP and consumption we use the level after detrending using the HP filter. For inflation, nominal wage inflation, the FF rate, and the 10-year U.S. Treasury rate, we demean by subtracting the period average. For the Fed's holdings of U.S. Treasuries, we use the change in the level from 2003, when data became available, as the effect of the Fed's purchases of government bonds. For deposits, we divide by the trend component of the GDP series to create a series relative to GDP, and then use this series after demeaning it by the period average.

Table C.1: Variable detail and source for macro estimation

Variable	Detail	Source
CPI, π	CPI for All Urban Consumers, All Items in U.S. City Average	FRED(CPIAUCSL)
Wage, π^w	Personal Income	FRED(PI)
Real GDP, y	Real Gross Domestic Product	FRED(GDPC1)
Consumption, c	Personal Consumption Expenditures	FRED(PCE)
Short-term Rate, r^b	Effective Federal Funds Rate	FRED(FF)
Long-term Rate, r^{bl}	U.S. Treasury Securities 10-Year	FRED(DGS10)
Central bank holdings, cb	Assets: Securities Held Outright: U.S. Treasury Securities	FRED(TREAST)
Loan Rate, r^l	Calculated from individual bank's data	See Table B.1
Deposit, D	Deposits, All Commercial Banks	FRED (DPSACBW027SBOG)

C.2.2 Estimation results

Table C.2 summarizes our estimation results, which include both priors and posteriors of estimated parameters. Given these estimated parameters, we obtain the time series of the estimated shocks. By using these shocks, we can see the contributions of each shock on the observable variables, which we show in Figure C.2.

Table C.2: Estimated parameters

		prior			posterior		
Description	parameter	Dist.	mean	std	mean	5%	95%
std. of monetary policy	σ_{mp}	IG	0.01	0.05	0.0064	0.0063	0.0065
std. of productivity	σ_a	IG	0.01	0.05	0.0144	0.0125	0.0165
std. of market risk taking	σ_{mrt}	IG	0.01	0.05	0.0052	0.0051	0.0052
std. of preference	σ_p	IG	0.01	0.05	0.0172	0.0168	0.0175
std. of markup	σ_μ	IG	0.01	0.05	0.0538	0.0486	0.0592
std. of deposit	σ_d	IG	0.01	0.05	0.4541	0.4122	0.4971
std. of risk-taking	σ_{rt}	IG	0.01	0.05	1.0013	0.9737	1.0294
std. of wage markup	σ_{μ^w}	IG	0.01	0.05	2.9759	1.2312	6.4005
std. of QE	σ_{qe}	IG	0.01	0.05	0.2627	0.2532	0.2751
AR of monetary policy	ρ_{mp}	Beta	0.5	0.2	0.0700	0.0622	0.0862
AR of productivity	ρ_a	Beta	0.5	0.2	0.8560	0.8270	0.8853
AR of market risk taking	ρ_{mrt}	Beta	0.5	0.2	0.0115	0.0028	0.0244
AR of preference	ρ_p	Beta	0.5	0.2	0.6285	0.5737	0.6699
AR of markup	ρ_μ	Beta	0.5	0.2	0.5481	0.5061	0.5914
AR of deposit	ρ_d	Beta	0.5	0.2	0.8540	0.8375	0.8703
AR of risk-taking	ρ_{rt}	Beta	0.5	0.2	0.7680	0.7368	0.8025
AR of wage markup	ρ_{μ^w}	Beta	0.5	0.2	0.6423	0.5787	0.7005
AR of QE	ρ_{qe}	Beta	0.5	0.2	0.5488	0.5212	0.5702
MA of markup	θ_μ	Beta	0.5	0.2	0.1129	0.0330	0.1926
MA of wage markup	θ_{μ^w}	Beta	0.5	0.2	0.4836	0.4287	0.5530
Capital adjustment cost	κ^i	Normal	5	1	0.0794	0.0653	0.0955
Wage adjustment cost	κ^w	Normal	0.1	0.1	0.0007	0.0002	0.0013
Price adjustment cost	κ^p	Normal	30	5	34.5709	34.4451	34.6846
Elasticity of LT bonds to ST bonds	ι^{st}	Normal	0	0.5	0.0035	0.0002	0.0104

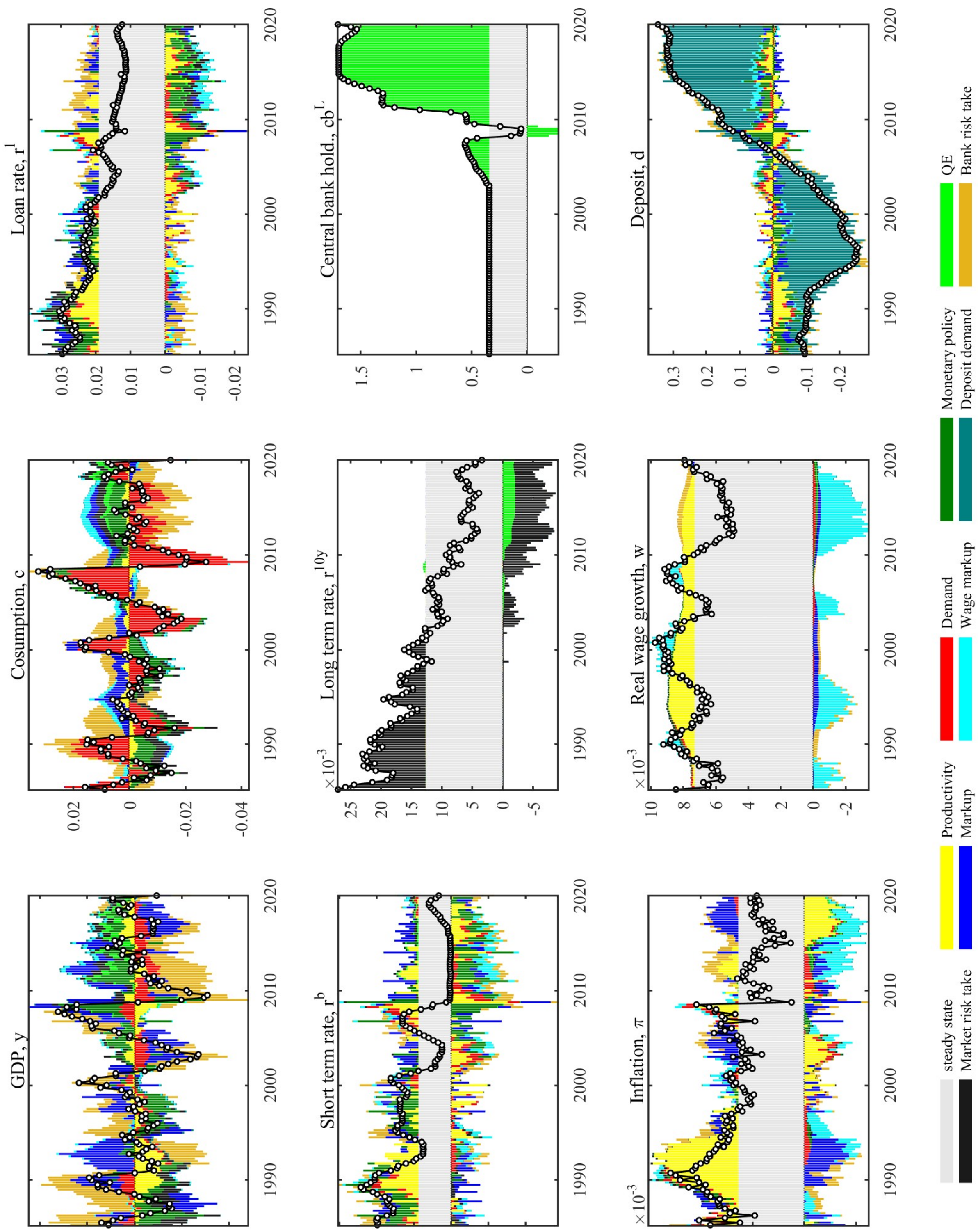

Figure C.2: Historical decompositions for each variable