Title
Essays on Banking and Monetary Economics

Author
Zhang, Mengbo

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays on Banking and Monetary Economics

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Economics

by

Mengbo Zhang

2021

ABSTRACT OF THE DISSERTATION

Essays on Banking and Monetary Economics

by

Mengbo Zhang
Doctor of Philosophy in Economics
University of California, Los Angeles, 2021
Professor Pierre-Olivier Weill, Chair

This dissertation consists of three chapters on banking and monetary economics. In Chapter 1, I study whether monetary policy is less effective in a low interest-rate environment. To answer this question, I examine how the passthrough of monetary policy to banks' deposit rates has changed, during the secular decline in interest rates in the U.S. over the last decades. In the data, the passthrough increased for about one third of banks, and decreased for the rest. Moreover, the deposit-weighted bank-average passthrough increased under a lower interest rate. I explain this observation in a model where banks have market power over loans and face capital constraints. In the model, when interest rates are low, the passthrough falls as policy rates fall, only in markets where loan competition is high. Hence, the overall passthrough depends on the distribution of loan market power. I confirm the model's prediction using branch-level data of U.S. banks. This channel also impacts the transmission of monetary policy to bank lending under low interest rates.

In Chapter 2 (joint with Tsz-Nga Wong), we document a new channel mediating the effects of monetary policy and regulation, the disintermediation channel. When the interest rate on excess reserves (IOER) increases, fewer banks are intermediating in the Fed funds

market, and they intermediate less. Thus, the total Fed funds traded decreases. Similarly, disintermediation happens after the balance sheet cost rises, e.g. the introduction of Basel III regulations. The disintermediation channel is significant and supported by empirical evidence on U.S. banks. To explain this channel, we develop a continuous-time search-and-bargaining model of divisible funds and endogenous search intensity that includes the matching model (e.g. Afonso and Lagos, 2015b) and the transaction cost model (e.g. Hamilton, 1996) as special cases. We solve the equilibrium in closed form, derive the dynamic distributions of trades and Fed fund rates, and the stopping times of entry and exit from the Fed fund market. IOER reduces the spread of marginal value of holding reserves, and hence the gain of intermediation. In general, the equilibrium is constrained inefficient, as banks intermediate too much.

In Chapter 3 (joint with Saki Bigio and Eduardo Zilberman), we compare the advantages of lump-sum transfers versus a credit policy in response to the Covid-19 crisis. The Covid-19 crisis has lead to a reduction in the demand and supply of sectors that produce goods that need social interaction to be produced or consumed. We interpret the Covid-19 shock as a shock that reduces utility stemming from "social" goods in a two-sector economy with incomplete markets. For the same path of government debt, transfers are preferable when debt limits are tight, whereas credit policy is preferable when they are slack. A credit policy has the advantage of targeting fiscal resources toward agents that matter most for stabilizing demand. We illustrate this result with a calibrated model. We discuss various shortcomings and possible extensions to the model.

The dissertation of Mengbo Zhang is approved.

Andrea Lynn Eisfeldt

Matthew Saki Bigio Luks

Andrew Granger Atkeson

Pierre-Olivier Weill, Committee Chair

University of California, Los Angeles

2021

To my parents

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am deeply grateful to my advisor, Pierre-Olivier Weill for his continuous support and guidance, and for his detailed and insightful comments on my dissertation. I am also indebted to my committee members: Andrew Atkeson, Saki Bigio, and Andrea Eisfeldt for their continuous supports, encouragement and fruitful discussions throughout the development of my dissertation. I am also thankful to my co-authors, Tsz-Nga Wong, Saki Bigio and Eduardo Zilberman, for their invaluable contributions. I would also like to thank, for great discussions and suggestions, Paula Beltran, Huifeng Chang, Yann Koby, Lee Ohanian, Facundo Piguillem, Shihan Shen, Liyan Shi, Jonathan Vogel, Stephen Williamson, Randall Wright, Shengxing Zhang, Feng Zhu, Yu Zhu. I gratefully acknowledge the support of the Lewis L. Clark Graduate Fellowship. All errors are my own.

VITA

Education	University of California, Los Angeles	Los Angeles, CA, USA
	C.Phil., Economics	2017
	M.A., Economics	2017
	Renmin University of China	Beijing, China
	M.A., Quantitative Economics	2017
	Zhejiang University	Hangzhou, Zhejiang, China
	B.A., Economics	2013
Experience	Federal Reserve Bank of Richmond	Richmond, VA, USA
	Dissertation Intern	2018
Award	UCLA Dissertation Year Fellowship	2020

CHAPTER 1

Loan Market Power and Monetary Policy Passthrough under Low Interest Rates

1.1 Introduction

In recent decades, advanced economies have experienced a secular decline in nominal interest rates. It is thought that this environment will persist as policy makers renew their efforts to stimulate these economies. However, previous research has observed that bank profitability may fall in a low-interest-rate environment (e.g., Jackson, 2015; Bech and Malkhozov, 2016; Claessens et al., 2018). This observation is a source of concern because it is thought that low bank profitability may actually impair banks' intermediation between deposits and loans (Bindseil, 2018). This impact can further weaken the transmission of monetary policy, which is usually measured as a lower passthrough from monetary policy rates to bank deposit rates, i.e., the extent to which commercial banks increase deposit rates in response to an increase in the monetary policy rate.[1] Motivated by this concern, this paper provides an empirical and theoretical investigation on how the passthrough of monetary policy rates to deposit rates changes as economies transition to a low-interest-rate environment.

On the empirical front, I study how the monetary policy passthrough changed for individual banks. I measure passthrough as the regression coefficient of the change in bank deposit rate on the monetary policy shocks. I document that as the U.S. transitioned to a low inter-

[1]For related discussion, see Brunnermeier and Koby (2018), Wang (2018), Balloch and Koby (2019), and Ulate (2021).

est rate environment, the monetary policy passthrough changed for most banks. However, the direction of change is not the same for every bank. For some banks the passthrough increased, whereas for others it decreased. This observation is important because it showcases that the concern that the overall passthrough of monetary policy will decline with lower rates depends on the distribution of individual bank's passthrough sensitivity to interest rate. Moreover, by taking into account the heterogeneous change in the passthrough across banks, I find that the aggregate passthrough to deposit rates is slightly higher under a lower interest rate in the U.S..

On the theoretical front, I build a model of bank competition to explain why the passthrough for banks changes differently as an economy transitions to a low-interest-rate environment. The model has the following features: First, the economy consists of a finite number of banks that raise deposits and invest the funds, together with equity, in loans and fixed-income bonds. Second, in both loan and deposit markets, banks engage in monopolistic competition by setting loan rates over loans, and deposit rates over deposits. For loan and deposit demand schedules, the elasticities of substitution across individual banks are both greater than one and the aggregate deposit demand has unit elasticity. Third, both banks and depositors can invest in bonds and earn a competitive rate of return set by the central bank, i.e., the Fed funds rate. The Fed funds rate thus also represents the nominal interest rate and equals the marginal cost of bank loans, and the difference between the Fed funds rate and deposit rate—the deposit spread—represents the cost of holding deposits. Finally, each bank is subject to a capital constraint, which is that a bank's deposit liabilities cannot exceed a multiple of bank profits. With these elements, I find that the change of a bank's passthrough under a lower interest rate depends on the degree of loan market concentration. As the interest rate falls, the monetary policy passthrough increases for banks located in concentrated loan markets, but decreases for banks located in competitive loan markets.

The model's mechanism works as follows. First, since banks can invest in bonds, the optimization on loans is independent of the deposit market, and the loan profit is a function

of the nominal interest rate and the number of banks in the loan market. Specifically, the loan profit is decreasing in the nominal rate, due to loan demand elasticity larger than one, and also decreasing in the number of banks in the loan market. More importantly, the loan profit decreases more in the nominal rate if the loan market is more concentrated, i.e., the loan market has fewer banks. On the other hand, the return on bonds is increasing in the nominal rate. This implies that a bank's profits on assets are non-monotonic in the nominal interest rate: the profit on loans is decreasing in the nominal interest rate, and the return on bonds is increasing in the nominal interest rate. Moreover, the decreasing effect on loan profits is stronger if the loan market is more concentrated.

Second, the non-monotonic effect of the nominal interest rate on asset profits is passed to the deposit side through the capital constraint. Due to the unit-elastic aggregate deposit demand, the capital constraint is always binding. This results in an equilibrium deposit spread that depends on a bank's profits on assets: When the nominal interest rate is low, for banks in concentrated loan markets, a lower interest rate induces a large increase in loan profits, which actually expands bank profits on assets. This allows banks to take more deposits at lower deposit spreads, thereby improving the passthrough. However, for banks in competitive loan markets, the profits on loans are not sensitive to the nominal rate. With a lower nominal rate, the profits on assets decrease due to the decline of return on bonds. As a consequence, banks have to take less deposits at higher deposit spreads, which weakens the passthrough.

The theoretical model has the following testable predictions. First, when the nominal interest rate is below (above) a threshold, the passthrough of monetary policy rates to deposit rates is increasing (decreasing) in the bank's loan market concentration. Second, as the interest rate falls, the passthrough increases if the bank's loan market concentration is sufficiently high, and decreases otherwise. The third testable prediction states that the threshold value of the nominal interest rate, which is mentioned in the first prediction, is decreasing in bank equity.

I test the model predictions using a panel dataset of U.S. bank branches. The main identification issue is that banks' lending and deposit opportunities simultaneously affect their decisions on deposit rates. In order to ensure that banks face similar deposit opportunities, I compare the deposit rates and deposit growth across branches within the same county, but belonging to different parent banks. This within-county estimation has two identifying assumptions. First, a bank can raise deposits at one branch and lend them at another, which implies that the impact of loan competition on a bank's deposit rates is determined by the average loan concentration of its branches. Second, bank competition over deposits and loans is localized in a banking market, and county is taken to be the unit of the banking market. As a result, the branches within the same county, but belonging to different banks, are faced with similar deposit opportunities and different loan market concentration.[2]

Moreover, the measurement of key variables is as follows. First, to account for the term premia, the measure of the nominal interest rate, which follows Wang (2018), is the yield rate of a treasury portfolio that replicates the repricing maturity of banks' loan portfolio. Second, the proxy for a bank's loan market concentration is the average Herfindahl index of home mortgage loans in counties, weighted by the bank's home mortgage lending across counties. Third, the proxy for a bank's equity is the ratio of total equity to total assets.

The empirical results support the theoretical predictions. First, when the nominal interest rate is zero and the Fed funds rate increases by 100 bps, banks that make loans in a monopoly market increase their deposit rates by 282 bps more than banks that make loans in a perfectly competitive market. At the same time, the corresponding growth of deposits is 266 bps larger at the banks that make loans in a monopoly market. This differential effect shrinks as the interest rate rises, and vanishes to zero when the nominal interest rate reaches a threshold. Second, as the interest rate falls, the passthrough to deposit rates increases if a bank's Herfindahl index of loans is higher than 0.21. Third, the threshold value of the nominal

[2]This identification assumption has been used in related empirical analysis on bank market power and passthrough, such as Drechsler et al. (2017), Drechsler et al. (2019), and Li et al. (2019).

interest rate, below which the passthrough increases in banks' loan market concentration, is decreasing in a bank's equity-assets ratio. For banks with average equity-assets ratio, the threshold value is between 2.5% and 3.0%, depending on the deposit product in regression.

I conduct several robustness tests of my findings. First, the results are robust if I control different fixed effects and bank characteristics. Second, the results are robust for large banks. Third, the results are robust for alternative measures of nominal interest rate, such as the Fed funds rate and 1-year Treasury yield rate. Fourth, I re-run the regressions with data before the global financial crisis of 2008, and obtain similar estimation results.

Next, I explore how this passthrough channel affects the response of banks' balance sheets to monetary policy. First, I verify that all of the branch-level regression results hold at the bank level using Call Reports data for U.S. banks. Second, I find that at a low interest rate, the impact of loan market power on monetary policy transmission to deposit growth also transmits to other balance sheet components: when the interest rate is low, an increase in the monetary policy rate induces higher growth of loans, securities, and assets, for banks with higher loan market concentration.

The above theoretical and empirical analysis shows that the level of nominal interest rates affects the monetary policy passthrough to deposit rates. However, my empirical analysis only identifies loan market competition as one determinant, but omits other potential determinants. To estimate the effects of a nominal interest rate on the passthrough comprehensively, I propose a new measure of monetary policy passthrough using Call Reports data. This measure is a pair of betas: zero beta and slope beta. For deposit rates, zero beta measures the passthrough of monetary policy rates to deposit rates when the nominal interest rate is zero; slope beta measures the change in the passthrough when the nominal rate increases by 100 bps. Hence, a negative (positive) slope beta means the passthrough increases (decreases) with a lower nominal interest rate.

The estimates suggest that the betas differ substantially across banks. First, the two betas are also highly negative correlated. This implies that banks with a high zero beta,

i.e., a high passthrough to deposit rates at a zero nominal rate, experience a decline in the passthrough as the nominal rate increases. However, for banks with a low zero beta at a zero nominal rate, passthrough actually increases with the nominal rates. This is consistent with the theoretical predictions. Second, the share of banks with a negative slope beta is 24.6%, implying that a low interest rate improves the passthrough for about about a quarter of banks in the sample. The share is even larger, 32.4%, for the largest 5% banks by assets. These numbers imply that whether the overall passthrough decreases with a lower interest rate depends on the distribution of market power. By weighting the slope betas by bank assets, I find that the average slope beta is about zero, implying that the overall passthrough does not significantly decrease with a lower interest rate. This is in contrast with the findings in the previous literature, which finds, by using models of representative banks, that a lower interest rate weakens passthrough.

To evaluate how the new measure of passthrough accounts for the impact of a nominal interest rate on monetary policy transmission, I estimate the analogous pair of betas for the growth of individual bank deposits, loans, securities, and assets. These betas measure the bank balance sheet's sensitivity to the policy rate at a zero nominal rate and the change of the sensitivity when the nominal rate increases by 100 bps. I find that the betas of deposit rates are significantly positive correlated with the betas of balance sheet growth. This implies that the impact of nominal rates on the passthrough to deposit rates can explain the effects of nominal interest rates on monetary policy transmission to bank balance sheets.

Related literature. This paper contributes to the literature on monetary policy transmission through banks.[3] Specifically, my paper highlights the role of bank market power

[3]Conventionally there is a large body of literaure on monetary policy transmssion through banks. The literature mostly focuses on two channels: the bank reserve channel and bank capital channel. In the bank reserve channel, monetary policy controls the size of bank reserves, which determines the size of bank deposits and hence bank lending. The related literature includes Bernanke (1983), Bernanke and Blinder (1988), Bernanke and Blinder (1992), Kashyap et al. (1993), Kashyap and Stein (1994), Kashyap and Stein (1995) Kashyap and Stein (2000), and Jiménez et al. (2014). The bank capital channel argues that a surprise rise in interest rate reduces bank assets by more than liabilities due to maturity mismatch, thus decreasing bank

as in Drechsler et al. (2017) and Scharfstein and Sunderam (2016), where Drechsler et al. (2017) demonstrate the role of deposits market power, and Scharfstein and Sunderam (2016) document the role of market power in loan markets. Unlike the previous literature, my work considers market power over deposits and loans simultaneously, and shows that the interplay between them reveals a new channel which impacts the response of bank deposits to monetary policy. In a recent paper, Wang et al. (2020) also study bank market power over deposits and loans jointly. But their paper focuses on comparing the quantitative effect of market power on monetary transmission versus the traditional channels, and does not discuss the theoretical implications of the interplay between deposit and loan competition.

Second, this paper relates to the recent literature on the effects of low interest rates (Krugman et al., 1998; Eggertsson and Woodford, 2006; Brunnermeier and Koby, 2018; Balloch and Koby, 2019; Ulate, 2021; Wang, 2018; Eggertsson et al., 2019; Bigio and Sannikov, 2021).[4] They argue that the usual transmission channels of monetary policy will be weakened or break down under very low interest rates, and short-run interest rate cuts or long-run low interest rate policy could be contractionary for aggregate bank lending. My paper differs from this research by focusing on the heterogeneous effects of low interest rates on the cross section of banks. I find that low interest rates impact banks heterogeneously: while a lower interest rate weakens the monetary policy passthrough for some banks, it can actually improve the passthrough for others. Moreover, a key determinant of the heterogeneity is banks' market power. This implies that the aggregate effect of low interest rates on monetary policy transmission really depends on the distribution of bank market power. In a related paper, Sá and Jorge (2019) investigate the validity of the deposit market power channel of monetary policy under low interest rates. They find that a low interest rate environment may turn

capital and compressing their lending capacity. The related literature includes Bemanke and Gertler (1989), Kiyotaki and Moore (1997), Gertler and Kiyotaki (2010), He and Krishnamurthy (2013), Brunnermeier and Sannikov (2014), Brunnermeier and Koby (2018), Van den Heuvel (2002), Bolton and Freixas (2000), and Brunnermeier and Sannikov (2016).

[4]Early research in this strand of literature focuses on the effects of a zero lower bound, while recent literature studies primarily the effects of negative interest rates.

this channel off.

Third, this paper is related to the literature studying the passthrough efficiency of monetary policy rates to deposit rates (Berger and Hannan, 1989; Hannan and Berger, 1991; Diebold and Sharpe, 1990; Neumark and Sharpe, 1992; Driscoll and Judson, 2013; Yankov, 2014; Drechsler et al., 2017; and Duffie and Krishnamurthy, 2016). This literature shows the adjustment of deposit rate to interest rate changes is slow and asymmetric, and interprets it as evidence of price rigidities or deposit market power. My work contributes to this literature by showing that the passthrough efficiency depends on the level of the nominal interest rate, and is heteregeneous across banks.

Finally, this paper contributes to the literature on the economic consequences of banking sector concentration. The effects of imperfect competition and market power in the banking sector, which arises from rising bank concentration, has been an active strand of research during past decades. Several papers find that the markups in the global and U.S. banking industry have risen greatly during the past four decades (De Loecker et al., 2020; Diez et al., 2018). Others provide quantitative analysis on the economic consequences of rising bank concentration and related regulations (Corbae and D'Erasmo, forthcoming). There are also papers that focus on the theoretical interactions between competition, financial fragility, and monetary policy (Corbae and Levine, 2019). My paper focuses on the role of loan market concentration on monetary policy passthrough, and provides new insights on how rising bank concentration would affect the real effects of monetary policy.

Outline. The remainder of the paper is as follows. Section 1.2 describes the motivating evidence. Section 1.3 presents the static model that rationalizes the motivating facts, as well as illustrates the key mechanism at play. Section 1.4 describes the data. Section 1.5 presents the empirical results that support the theoretical predictions, and proposes the new measure of monetary policy passthrough. The final section, 1.6, concludes the paper.

1.2 Motivating evidence

This section presents the empirical facts that motivate this paper. I find that for U.S. banks, a lower interest rate can change the passthrough to deposit rates heterogeneously. Specifically, when the interest rate is lower, some banks have a declined passthrough to deposit rates, while others have an increased passthrough to deposit rates.[5] Moreover, the aggregate passthrough to deposit rates is higher under a lower interest rate.

I start with the analysis by estimating the passthrough to a bank's deposit rate. Following Drechsler et al. (forthcoming), I use the balance sheet data for U.S. banks from the Call Reports between 2000Q1 and 2019Q4,[6] and run the following time-series regression for each bank j:

$$\Delta\text{Deposit Rate}_{j,t} = \alpha_j + \sum_{\tau=0}^{3} \delta_{j,\tau} \times \Delta i_{t-\tau} + \sum_{\tau=0}^{3} \beta_{j,\tau} \times i_{t-1-\tau} \times \Delta i_{t-\tau} + \theta_j \text{Controls}_{j,t-1} + \varepsilon_{j,t}, \quad (1.1)$$

where t is a quarter, ΔDeposit Rate$_{j,t}$ is the change in bank j's deposit rate from period $t-1$ to t, Δi_t is the monetary shock in period t, i_t is the level of interest rate in period t, and α_j are bank fixed effects. The deposit rate is the total quarterly interest expenses on domestic deposits divided by total domestic deposits and then annualized. It measures the average rate of interest expenses on deposits cumulated from the existing deposits. The monetary shock is the Fed funds rate shock from Nakamura and Steinsson (2018) cumulated at the quarterly level, and normalized to have a +100 bps impact on the interest rate i_t. I measure the interest rate i_t by the effective Fed funds rate in our baseline results. For

[5] As documented in previous literature, the aggregate passthrough of policy interest rates to deposit rates is incomplete in advanced economies (e.g. Drechsler et al., 2017, Wang, 2018, Balloch and Koby, 2019). That is, a 100 bps increase in the policy interest rate leads to less than 100 bps increase in aggregate bank deposit rate. As suggested by Drechsler et al. (2017), bank market power in the retail markets is a key driving force. In the presence of market power, an increase in the policy interest rate allows banks to raise the markups between the policy interest rate and deposit rate, since the opportunity cost of holding cash is higher.

[6] The details of data description are in Section 1.4.

robustness, I also use the shadow Fed funds rate from Wu and Xia (2016) as the alternative measure of i_t. I add three quarter lags of Δi_t and $i_{t-1} \times \Delta i_t$ to capture the cumulative effect of monetary policy shocks over one year. The variable $\text{Controls}_{j,t-1}$ represents the vector of control variables at bank level and aggregate level.[7]

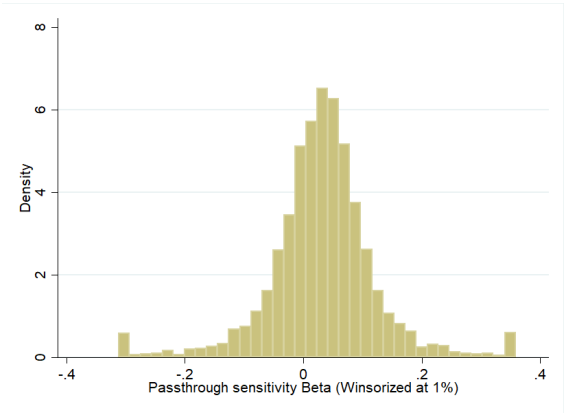
Note that the total impact of $\Delta i_{t-\tau}$ on $\Delta\text{Deposit Rate}_{j,t}$ is $\delta_{j,\tau} + \beta_{j,\tau} \times i_{t-1-\tau}$. It means that the passthrough to deposit rates is a linear function in the level of interest rate, and the sign of coefficient $\beta_{j,\tau}$ determines whether the passthrough is increasing or decreasing in the level of interest rate $i_{t-1-\tau}$. In this linear regression, I measure the passthrough of policy rates to bank deposit rates by $\sum_{\tau=0}^{3} \delta_{j,\tau} + \sum_{\tau=0}^{3} \beta_{j,\tau} \times i_{t-1-\tau}$, and measure the sensitivity of the passthrough to interest rate by $\beta_j \equiv \sum_{\tau=0}^{3} \beta_{j,\tau}$. The sensitivity estimate β_j represents the units of change in the passthrough if the interest rate is 1% higher over the past one year. The banks included are required to have at least 60 quarters of data in our sample period, which yields 4,596 banks.

Figure 1.1 shows the distribution of passthrough sensitivity β_j for all the banks. The left panel presents the distribution of estimates using the effective Fed funds rate as the measure of i_t. The right panel presents the distribution of estimates using the shadow Fed funds rate as the measure of i_t. The distribution of β estimates reports the following key messages. First, the estimates of passthrough sensitivity have a large variation across banks. For instance, in the left panel, about one third of banks have a negative passthrough sensitivity β, while the others have a positive β. A negative β means the passthrough to deposit rates increases under a lower interest rate, while a positive β means the passthrough decreases under a lower interest rate. Second, I measure the aggregate passthrough sensitivity by the average passthrough sensitivity β weighted by banks quarterly average inflation-adjusted deposits. I find that the average passthrough sensitivity is -0.006 if we use the effective Fed funds rate as the interest rate, and -0.04 if we use the shadow Fed funds rate as the interest rate. This

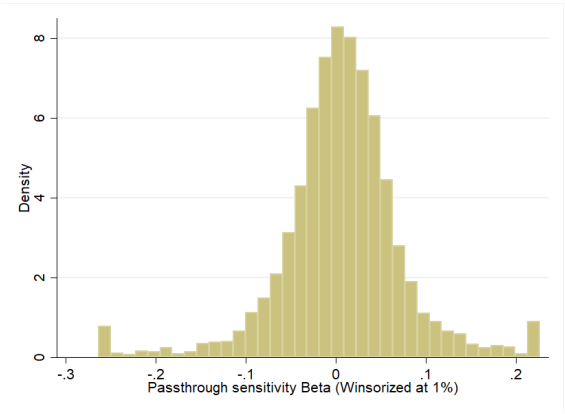
[7]The list of control variables include: four quarter lags of i_t, unemployment rate, real GDP growth rate, linear and quadratic time trend and a vector of bank-level controls. The bank-level controls include log assets, the share of core deposits in bank liabilities, loan-assets ratio and equity-assets ratio.

Figure 1.1: Distribution of passthrouth sensitivity β

(a) Effective Fed funds rate



(b) Shadow Fed funds rate



Notes: This figure plots the histogram of passthrough sensitivity β. The passthrough sensitivity β is estimated in equation (1.1). The left panel reports the estimates using the effective Fed funds rate as the measure of interest rate. The right panel reports the estimates using the shadow Fed funds rate from Wu and Xia (2016) as the measure of interest rate. Only banks with at least 60 quarterly observations are included. The estimates are winsorized at the 1% level. The underlying data are from FRED, the Call Reports and Nakamura and Steinsson (2018). The sample period is from 2000Q1 to 2019Q4.

implies that by taking into account the heterogeneous sensitivity across banks, the overall sensitivity of passthrough is slightly negative. In other words, the aggregate passthrough is slightly higher under a lower interest rate. This estimation result is opposite to the existing concern.

1.3 Baseline model

In this section I build a partial equilibrium model to study the impact of bank market power on the passthrough of Fed funds rates to deposit rates. The model is simplified to capture only the main economic force and allows for an analytical solution.

The economy lasts for one period and consists of three types of agents: (i) a unit mass of representative savers that demand deposits for transaction services; (ii) a mass μ of repre-

sentative borrowers that demand loans for consuming final goods; (iii) N banks of mass $1/N$ that engage in monopolistic competition over loans and deposits by setting the loan rates and deposit rates.[8] At the same time, the savers and banks can invest in or borrow from a class of fixed-income assets, which are called bonds. The bonds do not provide liquidity convenience and are traded in a competitive market with a common rate of return, which is exogenous and set by the central bank monetary policy. I refer to the rate of return as the Fed funds rate and denote it by i.[9] Moreover, I set the final goods consumed by borrowers as the numeraire. The deposits, loans and bonds are all measured in the units of final goods.

1.3.1 Demand for deposits

The demand for deposits is derived from the savers' utility maximization problem. The representative saver's utility function is

$$u^s(y, D) = y + \theta_d \cdot \ln(D), \tag{1.2}$$

where y is the consumption of final goods and D is the aggregate deposits that represents transaction services from deposit holdings.[10] The deposit aggregate D is a CES aggregate

[8] Assuming the mass of a bank to be $1/N$ is to guarantee that increasing the number of banks does not affect the equilibrium outcomes by mechanically increasing the total volume of loans and deposits, but through the extent of bank competition. This setup follows Drechsler et al. (2017) and Li et al. (2019).

[9] The maturity structure is not important for our main theoretical results. Thus we can think of the bonds as the Fed funds plus 3-month treasury bills, and the rate of return as the Fed funds rate that is used by the Federal Reserve to influence bank lending and deposit creation.

[10] See Section 1.3.4 for the interpretation of demand for deposits.

of bank deposits D_j from bank $j \in \{1, 2, ..., N\}$ with elasticity of substitution $\sigma_d > 1$:[11]

$$D = \left(\frac{1}{N} \sum_{j=1}^{N} D_j^{\frac{\sigma_d - 1}{\sigma_d}} \right)^{\frac{\sigma_d}{\sigma_d - 1}}. \tag{1.3}$$

I denote the deposit interest rate paid by bank j as i_j^d and let $s_j^d \equiv i - i_j^d$ be the spread between the Fed funds rate and the deposit rate. Since the household can invest in bonds, deposit spread is the opportunity cost of holding deposits.

The representative household is initially endowed with final goods y_0. It chooses the deposit holdings $\{D_j\}_{j=1}^N$ to maximize the utility function (1.2) subject to the budget constraint

$$y = (1+i) \left(y_0 - \frac{1}{N} \sum_{j=1}^{N} D_j \right) + \frac{1}{N} \sum_{j=1}^{N} \left(1 + i_j^d\right) D_j = (1+i)\, y_0 - \frac{1}{N} \sum_{j=1}^{N} s_j^d D_j, \tag{1.4}$$

where I replace i_j^d with $i - s_j^d$ in the second equality. Thus the deposit spread s_j^d is the price of bank j's deposits in the budget constraint.

Note that the household's problem is equivalent to the following two-step problem. First, since the CES aggregator of deposits (1.3) is constant returns to scale, the household selects the deposit holdings across banks to minimize the unit cost of aggregate deposits. The

[11]Note that D_j formally represents the amount of deposits when the mass of bank j is scaled to one. Since the mass of a bank is $1/N$, the real amount of bank j's deposits held by the household is $\frac{1}{N} D_j$. In the rest part of the model, the amounts of loans, bonds and equity of an individual bank are defined in an analogous way.

minimized unit cost is the average deposit spread given by[12]

$$s^d \equiv \min_{\{D_j\}_{j=1}^N} \frac{1}{N}\sum_{j=1}^{N} s_j^d D_j \text{ s.t. } D\left(\{D_j\}_{j=1}^N\right) = 1 \tag{1.5}$$

$$= \left[\frac{1}{N}\sum_{j=1}^{N}\left(s_j^d\right)^{1-\sigma_d}\right]^{\frac{1}{1-\sigma_d}}, \tag{1.6}$$

where $D\left(\{D_j\}_{j=1}^N\right)$ denotes the deposit aggregator (1.3). Second, the representative household maximizes the utility over consumption and aggregate deposits according to the following problem:

$$\max_{D,y} y + \theta_d \cdot \ln(D) \tag{1.7}$$

subject to

$$y = (1+i)\, y_0 - s^d \cdot D. \tag{1.8}$$

Solving the above problem (1.5) and (1.7), I obtain that the demand for aggregate deposits is a function of the average deposit spread:

$$D\left(s^d\right) = \frac{\theta_d}{s^d}, \tag{1.9}$$

and the demand for bank j's deposits is

$$D_j\left(s_j^d; s_{-j}^d\right) = \left(\frac{s^d}{s_j^d}\right)^{\sigma_d} D\left(s^d\right), \tag{1.10}$$

where s_{-j}^d denotes the deposit spread of all banks except j. Note that the competitors' deposit spread determines bank j's deposit demand since the average deposit spread s^d, which is given by (1.6), is a function of $\left\{s_j^d\right\}_{j=1}^N$. Thus a change in s_j^d affects not only the individual demand D_j, but also the aggregate demand D through changing s^d. Moreover,

[12]The derivations for the baseline model are provided in the Appendix 1.A.1.

the demand elasticity for individual bank deposits is:

$$\epsilon_j^d = -\frac{\partial \log\left(D_j\right)}{\partial \log\left(s_j^d\right)} = \sigma_d + \frac{1}{N}\left(1-\sigma_d\right)\left(\frac{s_j^d}{s^d}\right)^{1-\sigma_d}. \tag{1.11}$$

Note that the individual demand elasticity is increasing in N due to $\sigma_d > 1$. This implies that the individual deposit demand is more elastic if there are more banks in the deposit market. Moreover, if all the banks set the same deposit rate, the individual deposit spread s_j^d is equal to the average deposit spread s^d. Then the demand for individual deposits becomes identical, i.e. $D_j = D\left(s^d\right)$, and the individual demand elasticity becomes a constant, i.e. $\epsilon_j^d = \epsilon^d\left(N\right) \equiv \sigma_d + \frac{1}{N}\left(1-\sigma_d\right) > 1$.

1.3.2 Demand for loans

The demand for loans is derived from the borrowers' utility maximization problem. The representative borrowers have a quasi-linear utility over final goods consumption and labor supply:

$$u^b\left(c^b, h\right) = \frac{\left(c^b\right)^{1-\nu} - 1}{1-\nu} - \theta_h \cdot h, \tag{1.12}$$

where c^b represents the consumption of final goods, h represents the amount of labor supply. The parameter θ_h the disutility of labor supply, and $\nu < 1$ is the parameter of CRRA utility. The borrowers have access to a production technology that produces one unit of final goods for each unit of labor supply. However, the demand for final goods consumption arrives before the production, and the borrowers have no initial endowment.[13] Then the consumption demand must be financed by the aggregate bank loans L, which is a CES

[13]This is a simple way to model the demand for loans. See Section 1.3.4 for detailed discussions of the assumption.

aggregate of individual bank loans L_j with the elasticity of substitution $\sigma_l > \frac{1}{\nu}$:

$$L = \left(\frac{1}{N} \sum_{j=1}^{N} L_j^{\frac{\sigma_l - 1}{\sigma_l}} \right)^{\frac{\sigma_l}{\sigma_l - 1}}. \tag{1.13}$$

Thus the representative borrowers' optimization problem is to maximize (1.12) subject to (1.13) and

$$c^b \leq L \text{ and } \frac{1}{N} \sum_{j=1}^{N} \left(1 + i_j^l\right) L_j \leq h, \tag{1.14}$$

where i_j^l is bank j's loan interest rate. Similar to the savers' problem, the borrowers' problem can also be solved in two steps. First, since the loan aggregator (1.13) is constant returns to scale, the borrowers distribute loans across banks to minimize the unit cost of aggregate loans. The minimized unit cost is the average loan rate i^l that is given by

$$1 + i^l \equiv \min_{\{L_j\}_{j=1}^N} \frac{1}{N} \sum_{j=1}^{N} \left(1 + i_j^l\right) L_j \text{ s.t. } L\left(\{L_j\}_{j=1}^N\right) = 1 \tag{1.15}$$

$$= \left[\frac{1}{N} \sum_{j=1}^{N} \left(1 + i_j^l\right)^{1-\sigma_l} \right]^{\frac{1}{1-\sigma_l}}. \tag{1.16}$$

Second, the buject constraints in (1.14) must be binding at optimum. Thus the borrowers maximize the utility by solving

$$\max_L \frac{L^{1-\nu} - 1}{1 - \nu} - \theta_h \cdot \left(1 + i^l\right) L. \tag{1.17}$$

The optimal solution to the demand for aggregate loans and individual bank loans are given by

$$L\left(i^l\right) = \mu \cdot \theta_h^{-\frac{1}{\nu}} \left(1 + i^l\right)^{-\frac{1}{\nu}}, \tag{1.18}$$

where μ is the mass of borrowers, and

$$L_j\left(i_j^l; i_{-j}^l\right) = \left(\frac{1+i^l}{1+i_j^l}\right)^{\sigma_l} L\left(i^l\right). \tag{1.19}$$

Moreover, the demand elasticity of individual bank loans is

$$\epsilon_j^l = -\frac{\partial \log\left(L_j\right)}{\partial \log\left(1+i_j^l\right)} = \sigma_l + \frac{1}{N}\left(\frac{1}{\nu} - \sigma_l\right)\left(\frac{1+i_j^l}{1+i^l}\right)^{1-\sigma_l}. \tag{1.20}$$

Note that the demand elasticity is increasing in N due to $\sigma_l > \frac{1}{\nu}$. With a higher N, the loan market is more competitive, and the loan demand curve of an individual bank becomes more elastic. Moreover, if all banks sent the same loan rate, the demand for individual bank loans is identical, i.e. $L_j = L\left(i^l\right)$, and the individual demand elasticity becomes $\epsilon_j^l = \epsilon^l(N) \equiv \sigma_l + \frac{1}{N}\left(\frac{1}{\nu} - \sigma_l\right) > 1$.

1.3.3 Bank's problem

Each bank is endowed with an identical equity E_0, which is a constant and independent of the Fed funds rate.[14] The banks raise deposits D_j, and invests the funds in loans L_j and bonds A_j. Thus the balance sheet identity is given by

$$L_j + A_j = D_j + E_0, \tag{1.21}$$

and Table 1.1 displays the structure of a bank balance sheet.

Moreover, each bank is subject to a capital constraint that restricts the amount of assets

[14]We can assume the equity is decreasing in the Fed funds rate due to maturity mismatch. However, our results do not change as long as $|\partial E_0/\partial i|$ is bounded.

Table 1.1: Bank balance sheet

Assets	Liabilities
Loans (L)	Deposits (D)
Bonds (A)	Equity (E_0)

it can hold on the balance sheet:

$$\psi \left(L_j + A_j \right) \leq N_j. \tag{1.22}$$

The parameter ψ is the risk weight and N_j denotes bank j's net worth at the end of the period:

$$N_j = \left(1 + i_j^l \right) L_j + \left(1 + i \right) A_j - \left(1 + i_j^d \right) D_j. \tag{1.23}$$

Taking the individual demand function for loans and deposits as given, banks simultaneously set loan rates and deposit spreads to maximize their own net worth. By using the balance sheet identity (1.21), we can replace A_j with $D_j + E_0 - L_j$ and write an individual bank's problem as

$$\max_{i_j^l, s_j^d} N_j = \left(i_j^l - i \right) L_j \left(i_j^l; i_{-j}^l \right) + s_j^d D_j \left(s_j^d; s_{-j}^d \right) + \left(1 + i \right) E_0 \tag{1.24}$$

subject to

$$\psi \left[D_j \left(s_j^d; s_{-j}^d \right) + E_0 \right] \leq N_j, \tag{1.25}$$

where $\left(i_j^l - i \right) L_j$ is the profit of loans, $s_j^d D_j$ is the profit of deposits, and $\left(1 + i \right) E_0$ is the gross gain of investing equity in bonds.

Two-step problem. An important feature of banks' problem is that banks' optimal decisions of loan rates are independent of the deposit market and the capital constraint. To see this, note that banks can substitute loans for bonds without changing the total amount of assets. This implies that the marginal funding cost for loans is the Fed funds rate i, even

though a bank can raise deposits. Therefore, we can solve the equilibrium in two steps. First, banks simultaneously choose loan rates to maximize their profits on loans:

$$\Pi_j^l = \max_{i_j^l} \left(i_j^l - i\right) L_j \left(i_j^l; i_{-j}^l\right), \; j = 1, 2, ..., N \tag{1.26}$$

Without loss of generality we focus on the equilibria where the individual loan rate i_j^l is non-negative. The following lemma characterizes the loan problem solution:

Lemma 1.1 *There exists a unique equilibrium to the loan problem. In this equilibrium, banks set identical loan rates*

$$1 + i_j^l = 1 + i^l = \frac{\epsilon^l(N)}{\epsilon^l(N) - 1}(1 + i), \tag{1.27}$$

and earn identical loan profits

$$\Pi_j^l = \Pi^l(i, N) = \mathcal{C}_l(N) \cdot (1 + i)^{-\frac{1-\nu}{\nu}}, \tag{1.28}$$

where

$$\mathcal{C}_l(N) \equiv \mu \cdot \theta_h^{-\frac{1}{\nu}} \frac{\left[\epsilon^l(N) - 1\right]^{\frac{1-\nu}{\nu}}}{\left[\epsilon^l(N)\right]^{\frac{1}{\nu}}} \tag{1.29}$$

is a decreasing function in N.

This lemma shows that due to the symmetry of individual bank loans in the loan aggregator (1.13), the symmetric equilibrium is the unique equilibrium of banks' loan problem. The equilibrium loan rate is a markup over the Fed funds rate, where the markup is a function of demand elasticity $\epsilon^l(N)$. The markup is decreasing in N since $\epsilon^l(N)$ is increasing in N. In the loan profit function $\Pi^l(i, N)$, the function $\mathcal{C}_l(N)$ is defined as the loan concentration index, which decreases with the number of banks N and the elasticity of substitution σ_l. Thus the loan profit $\Pi^l(i, N)$ also decreases in N. Intuitively, a larger N means a more competitive loan market, thus banks are less able to charge markup and earn monopolistic

profits. On the other hand, $\Pi^l(i, N)$ is also decreasing in the interest rate i. This is because the individual loan demand elasticity is greater than 1, which implies that an increase in the loan rate induces larger reduction in the loan quantity. Importantly, loan profits decrease less in i for a higher N. This implies that a bank's profits on loans is less sensitive to nominal rate if the loan market is more competitive.

Another important implication of Lemma 1.1 is that bank's earnings on assets are non-monotonic in the Fed funds rate: the loan profits $\Pi^l(i, N)$ is decreasing in i, and the return on bonds $(1+i)A_j$ is increasing i. The non-monotonicity means that when the Fed funds rate is low (high), a bank's earnings on assets is decreasing (increasing) in the Fed funds rate. Moreover, the non-monotonicity depends on the degree of loan market competition: when the Fed funds rate is low, an increase in the Fed funds rate induces larger decrease in a bank's asset earnings if it locates in a more concentrated loan market. As is shown below, the non-monotonic effect is then passed to the liability side through the capital constraint, and explains the pattern of passthrough documented in the data.

The second step is to solve the problem of deposits. Given the equilibrium solution to the loan problem, banks simultaneously set deposit spreads to maximize their deposit profits:

$$\Pi_j^d = \max_{s_j^d} \left\{ s_j^d D_j \left(s_j^d; s_{-j}^d \right) \right\} \tag{1.30}$$

subject to

$$\psi D_j \left(s_j^d; s_{-j}^d \right) \leq W(i, N) + s_j^d D_j \left(s_j^d; s_{-j}^d \right) \tag{1.31}$$

where

$$W(i, N) = \Pi^l(i, N) + (1 + i - \psi) E_0. \tag{1.32}$$

The deposit problem shows that the capital constraint is actually a Kiyotaki-Moore constraint: a bank's liabilities cannot exceed a multiple of future profits to ensure incentive compatibility. The future profits consist of two parts: $W(i, N)$ represents the effective profits on

bank assets, and $s_j^d D_j\left(s_j^d; s_{-j}^d\right)$ represents the profits on deposits. The non-monotonic effect of nominal interest rate on banks' asset earnings impacts the equilibrium spread through changing the shadow cost of the capital constraint. Without loss of generality we focus on the equilibria with finite deposit spreads. The following lemma summarizes the equilibrium solution to the deposit spreads.

Lemma 1.2 *There exists a unique equilibrium to the deposit problem. In this equilibrium, the capital constraint is always binding, and banks set identical deposit spreads*

$$s_j^d = s^d\left(W\left(i, N\right)\right) = \frac{\psi\theta_d}{\mathcal{C}_l\left(N\right)\cdot\left(1+i\right)^{-\frac{1-\nu}{\nu}} + \left(1+i-\psi\right)E_0 + \theta_d} \tag{1.33}$$

and earn constant and identical deposit profits

$$\Pi_j^d = \theta_d. \tag{1.34}$$

Therefore, the bank's problem (1.24) and (1.25) has a unique equilibrium that is described by (1.27), (1.28), (1.33) and (1.34).

This lemma has the following results. First, since the aggregate demand of deposits has a unit elasticity, an equilibrium without capital constraint has zero deposit spreads. This implies banks' deposit liabilities are infinite. However, the unit demand elasticity also implies constant profits on deposits. Hence, the capital constraint must be binding in equilibrium, and the equilibrium deposit spread is an increasing function in the shadow cost of the constraint. This means that the aggregate supply of deposits by banks is given by the binding capital constraint, i.e.

$$D^s\left(s^d\right) = \frac{\mathcal{C}_l\left(N\right)\cdot\left(1+i\right)^{-\frac{1-\nu}{\nu}} + \left(1+i-\psi\right)E_0}{\psi - s^d}, \tag{1.35}$$

where the average spread s^d represents the price of deposits. Then the equilibrium average

deposit spread is given by the equating the deposit supply and the deposit demand $D^d\left(s^d\right) = \frac{\theta_d}{s^d}$. Another important implication of this result is that the passthrough is one if the capital constraint is non-binding, and any incomplete passthrough in this model must come from the binding constraint.

Second, the equilibrium deposit spreads $s^d\left(W\left(i,N\right)\right)$ is a decreasing function in asset earnings $W\left(i,N\right)$. This is intuitive because a higher $W\left(i,N\right)$ makes the capital constraint less binding, thereby lowers the shadow cost of the capital constraint.

1.3.4 Discussion of model assumptions

Demand for liquidity. Incorporating liquidity services in the utility is the simplest way to generate a demand for assets with dominated return, such as deposits. An interpretation of this setup is that transactions are subject to frictions, such as cash in advance or frictions that arise from anonymous transaction and lack of commitment. Alternative ways would be to explicitly model these frictions, but models with these frictions produce similar utility functions as assuming liquidity in utility.[15] Moreover, the imperfect substitution between individual bank deposits captures the heterogeneous proximity, switching costs, tastes, or asymmetric information for depositors.

In general, modelling liquidity in utility usually incorporates currency and deposits as two substitutable forms of liquidity through an aggregator (Chetty 1969; Poterba and Rotemberg 1987; Nagel 2016; Di Tella and Kurlat forthcoming; Drechsler et al. 2017). However, the liquidity demand for currency is not necessary in this model. The main trade-off is the substitution between deposits and bonds, generating a comovement between the Fed funds rate and deposit rate. Thus in the baselien model I drop currency in utility to eliminate the impact of deposit competition on passthrough, which allows for an analytic solution for evaluating the impact of loan competition on passthrough.

[15]See, for example, Feenstra (1986) and Williamson (2012).

Demand for loans. In this model, the demand of loans by borrowers arises from the arrival of consumption demand prior to the time when borrowers earn income. One can build a richer model with two subperiods. Borrowers prefer to consume more in the first subperiod, but have more income in the second period. Thus borrowers use bank loans to smooth the consumption across two subperiods. For individual banks, each bank produces a differentiated loan product that matures in one period. The heterogeneity of bank loans is motivated by factors such as geographic location and industry expertise. The maturity of loans will play no particular role in the qualitative mechanism of this model. Ulate (2021) provides a microfoundation of loan demand that nests the above characteristics, and it can also be applied in my model.

A decreasing loan profit in nominal interest rate is the key feature of this model. The related empirical evidence has been documented in Scharfstein and Sunderam (2016) and Fuster et al. (2021). This feature generates a non-monotonic effect of interest rate on bank profitability, which is able to explain the impact of low interest rate on passthrough.

Capital constraint. It is useful to note that the capital constraint arises from multiple reasons. First, banks are subject to regulatory constraints, such as Basel III, which requires sufficient bank equity to support lending. Our constraint, which is based on Brunnermeier and Koby (2018), is a simple form that delivers the same requirement on bank balance sheet.[16] We can also motivate the constraint with multiple microfoundations that reflects banks' endogenous risk-taking behavior and agency problems, such as Holmstrom and Tirole (1997) and Gertler and Kiyotaki (2010). Second, the constraint can be replaced with a smooth convex leverage cost, such as Piazzesi and Schneider (2018).

To present the simplest model without losing the main mechanism, we do not introduce any other balance sheet frictions which are commonly used in other banking models, such as the reserve constraint, liquidity coverage constraint or the cost of non-reservable borrowing.

[16]For alternative forms of capital constraints, see, for example, Begenau et al. (2020) and Wang (2018).

These frictions do not change the qualitative results of the model.

1.3.5 Impacts of loan market power on passthrough

This section presents the comparative statics of loan market power and nominal interest rate on the passthrough of policy rate to deposit rate. Using the results of Lemma 1.2, the equilibrium deposit rate is given by

$$i^d(i,N) = i - s^d(W(i,N)). \tag{1.36}$$

Hence, the passthrough to deposit rates is defined as

$$\frac{\partial i^d(i,N)}{\partial i} = 1 - \frac{\partial s^d(W(i,N))}{\partial i} = 1 - \frac{\partial s^d(W(i,N))}{\partial W} \cdot \frac{\partial W(i,N)}{\partial i}. \tag{1.37}$$

Note that when the interest rate is low, the passthrough is incomplete, i.e. $\frac{\partial i^d(i,N)}{\partial i}$ is less than one, because $\frac{\partial s^d(W(i,N))}{\partial W} < 0$ and $\frac{\partial W(i,N)}{\partial i} < 0$. The following proposition describes the conditions needed for this result.

Proposition 1.1 *The passthrough of Fed funds rate to deposit rate* $\frac{\partial i^d}{\partial i}$ *is less than one iff* $i < i_0(\mathcal{C}_l, E_0)$*, where*

$$i_0(\mathcal{C}_l, E_0) \equiv \left(\frac{1-\nu}{\nu}\frac{\mathcal{C}_l}{E_0}\right)^{\nu} - 1. \tag{1.38}$$

This proposition states that the passthrough is incomplete when the nominal interest rate is below a threshold. This is because when the interest rate is low, an increase in interest rate depletes bank profit, thereby tightens the capital constraint and increases spread. Since an incomplete passthrough is consistent with the data of U.S. banks, we will focus on $i < i_0(\mathcal{C}_l, E_0)$ in the following analysis. This means all the propositions below are based on the following assumption, which guarantees a positive $i_0(\mathcal{C}_l, E_0)$:

Assumption 1.1 $\mathcal{C}_l > \frac{\nu}{1-\nu}E_0$.

As described in the following proposition, our first result is about the impact of loan market concentration on equilibrium deposit rate and passthrough: that, when the interest rate is sufficiently low, banks with higher market power set higher deposit rates than those in less concentrated markets.

Proposition 1.2 *(i) The equilibrium deposit rate increases in loan market concentration, i.e.*

$$\frac{\partial i^d}{\partial \mathcal{C}_l} > 0. \tag{1.39}$$

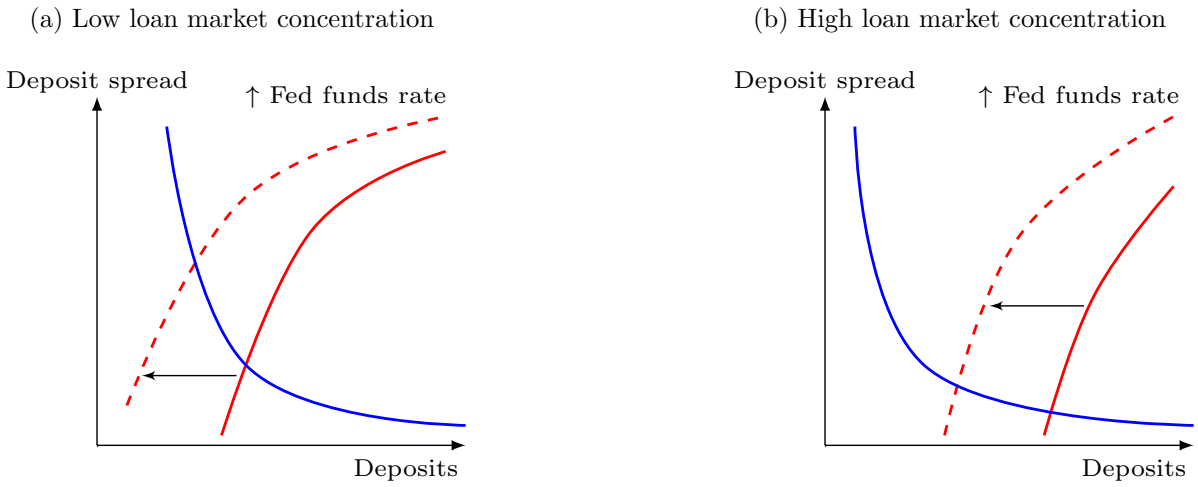
(ii) Given $\mathcal{C}_l$ and E_0, there exists a unique threshold $i_1(\mathcal{C}_l, E_0)$, which is smaller than $i_0(\mathcal{C}_l, E_0)$, such that the passthrough of Fed funds rate to deposit rate increases with loan market concentration iff $i < i_1(\mathcal{C}_l, E_0)$. That is,

$$\frac{\partial^2 i^d}{\partial i \partial \mathcal{C}_l} > (<)\, 0 \text{ iff } i < (>)\, i_1(\mathcal{C}_l, E_0). \tag{1.40}$$

(iii) The threshold $i_1(\mathcal{C}_l, E_0)$ is increasing in $\mathcal{C}_l$ and decreasing in E_0. Moreover, it is positive iff $\mathcal{C}_l > \theta_d + \frac{1+\nu}{1-\nu} E_0$.

It is useful to illustrate the intuition of Proposition 1.2 in graphics, as plotted in Figure 1.2. This figure plots the demand (blue) and supply (red) curves of aggregate deposits for high loan concentration and low loan concentration. First, the deposit rate is increasing in loan concentration because the loan profits increases with loan concentration. This increases bank's total profit and makes the capital constraint less binding. As a result, banks are able to take more deposits at lower deposit spreads. The solid red curves in Figure 1.2 show this result. Second, when the interest rate is sufficiently low, the location of deposit supply curve is determined by bank's loan market concentration. In this case, the supply curve of banks with higher loan concentration interacts with the demand curve in the more elastic part. As a consequence, when the interest rate increases, the deposit spread increases less with a higher loan market concentration. Third, the threshold interest rate $i_1(\mathcal{C}_l, E_0)$ represents the

Figure 1.2: Graphic Representation of Passthrough under Different Loan Market Concentration



Notes: This figure plots the passthrough of Fed funds rate to deposit rate for banks with low and high loan market power. The red curves are the supply of deposits, the blue curves are the demand of deposits.

cutoff rate below which the effect of loan concentration on passthrough dominates. A higher loan concentration implies a stronger effect of loan concentration on passthrough, while a higher equity implies a stronger effect of bond earnings on passthrough. This implies that the threshold value $i_1(\mathcal{C}_l, E_0)$ is increasing in $\mathcal{C}_l$ and decreasing in E_0.

Next we examine the net effect of low interest rate on passthrough efficiency through our loan market competition channel. We are interested in whether a lower interest rate improves or weakens a bank's passthrough to deposit rate, and its relationship with loan market power. The literature documents that low interest rate environment weakens the aggregate passthrough (e.g. Wang (2018)). However, the following proposition shows that the impact differs across banks. Specifically, a lower interest rate strengthens the deposit rate passthrough for banks with sufficiently large loan market concentration, but weakens otherwise.

Proposition 1.3 *Suppose* $\nu < \frac{1}{2}$. *(i) Given* $\mathcal{C}_l$ *and* E_0, *there exists a unique thresh-*

old $i_2(\mathcal{C}_l, E_0)$*, which is smaller than* $i_0(\mathcal{C}_l, E_0)$*, such that a lower interest rate improves passthrough if* $i < i_2(\mathcal{C}_l, E_0)$*, i.e.*

$$\begin{aligned} \frac{\partial^2 i^d}{\partial i^2} &< 0 \text{ if } i < i_2(\mathcal{C}_l, E_0), \\ \frac{\partial^2 i^d}{\partial i^2} &> 0 \text{ if } i \in (i_2(\mathcal{C}_l, E_0), i_0(\mathcal{C}_l, E_0)). \end{aligned}$$

(ii) Given E_0*, there exists a unique threshold* $\widehat{\mathcal{C}}_l(E_0) > 0$ *such that* $i_2(\mathcal{C}_l, E_0)$ *is positive iff* $\mathcal{C}_l > \widehat{\mathcal{C}}_l(E_0)$*.*

(iii) The threshold $\widehat{\mathcal{C}}_l(E_0)$ *is increasing in* E_0*; the threshold* $i_2(\mathcal{C}_l, E_0)$ *is increasing in* $\mathcal{C}_l$ *and decreasing in* E_0*.*

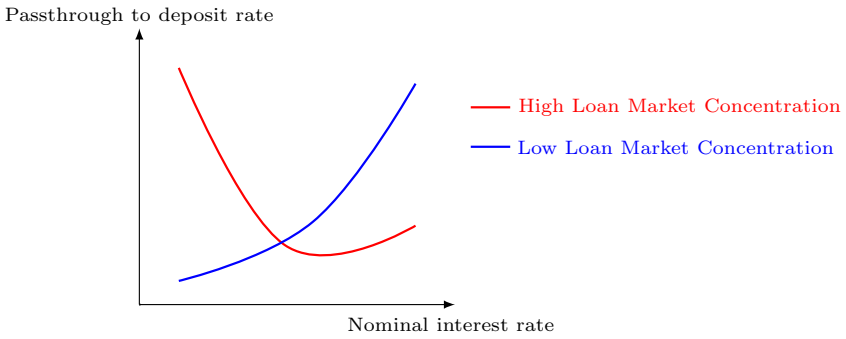
A graphic representation of Proposition 1.3 is plotted in Figure 1.3. The intuition is as follows. According to Proposition 1.2, a higher loan market concentration improves passthrough when the interest rate is low. This effect increases with a lower interest rate if and only if the sensitivity of loan profits to interest rates also increases with lower interest rate. This requires $\nu < \frac{1}{2}$ and a sufficiently large $\mathcal{C}_l$. The main implication of this proposition is that a lower interest rate can improve passthrough for banks with high loan concentration, while weakens passthrough if the loan market is competitive.

In sum, the above propositions imply the following testable predictions. First, the following equation summarizes the main results of Proposition 1.2:

$$\frac{\partial i^d}{\partial i} = \underset{+}{\beta_1} + \underset{-}{\beta_2} \times i + \underset{-}{\beta_3} \times E_0 \times \text{loan market concentration} + \text{other terms}, \tag{1.41}$$

As predicted by the propositions, we expect that $\beta_1 > 0$, $\beta_2 < 0$ and $\beta_3 < 0$. A positive β_1 represents that the passthrough to deposit rates increases in loan concentration under low interest rate. A negative β_2 implies that this relationship reverses under a high nominal rate. A negative β_3 means that with a higher bank equity, it is less likely that the deposit rate beta increases in bank concentration due to a more relaxed capital constraint. Note that the

Figure 1.3: Interest Rate Passthrough under Different Loan Market Concentration



Notes: This figure plots the theoretical prediction on the relationship between deposit spreads and nominal interest rates for banks with low loan market concentration (small $\mathcal{C}_l$) and high loan market concentration (large $\mathcal{C}_l$).

threshold value of interest rate $i_1(\mathcal{C}_l, E_0) = \frac{1}{-\beta_2}(\beta_1 + \beta_3 \times E_0)$ is decreasing in E_0, which is consistent with Proposition 1.2.

The second testable prediction is derived from Proposition 1.3:

$$\frac{\partial i^d}{\partial i} = \left(\underset{+}{\beta_4} + \underset{-}{\beta_5} \times \text{loan market concentration}\right) \times i + \text{other terms}, \tag{1.42}$$

We expect that $\beta_4 > 0$ and $\beta_5 < 0$. The signs of coefficients imply that if the loan market concentration is larger than $\frac{\beta_4}{-\beta_5}$, then the passthrough increases with a lower interest rate. Otherwise, the passthrough decreases with a lower interest rate.

1.4 Data

This section describes the data sources and the construction of main variables.

1.4.1 Data sources

Branch deposits. The data on branch-level deposit volumes are from the Federal Deposit Insurance Corporation (FDIC). The data cover the universe of U.S. bank branches at an annual frequency from June 1994 to June 2019. The information on branch characteristics, such as the parent bank, address, and geographic location, are also available. I use the FDIC branch identifier to match the FDIC data with other datasets.

Branch deposit rates. The data on retail interest rates are provided by Ratewatch. Ratewatch surveys bank branches across the U.S. and collects weekly branch-level deposit rates by products. I use the sample from January 2001 to December 2019. Compared with the branch information from the FDIC, the data cover 54% of all U.S. branches as of 2019. Following Drechsler et al. (2017), I restrict the data sample to the branches that actively set retail rates, which cover approximately 30% of all unique branches. Moreover, my analysis focuses on the following deposit products: 25K money market accounts (the money market deposit accounts with an account size of $25,000) and 10K CDs with 3-month, 6-month and 12-month maturities (the certificates of deposits with an account size of $10,000 and mature in 3 months, 6 months and 12 months). These products are commonly offered across U.S. branches, and are representative of savings and time deposit products. The Ratewatch data also report the FDIC branch identifier, thus I use it to match the Ratewatch data with the FDIC data.

Bank data. The bank data are from the U.S. Call Reports provided by the Federal Reserve Bank of Chicago and the Federal Financial Institutions Examination Council (FFIEC). The data contain quarterly income statements and balance sheets of all U.S. commercial banks. Our sample is from 1997Q1 to 2019Q4. I use the FDIC bank identifier to merge the bank-level Call Reports data with the branch-level FDIC and Ratewatch data.

Home mortgage loans. I collect the administrative data on residential mortgage loans from the Home Mortgage Disclosure Act (HMDA) dataset. The dataset covers the loan-level information on residential mortgages originated or purchased by most mortgage lending institutions in the U.S. at an annual frequency. In particular, it reports the amount of mortgage loans issued by a financial institution in a given county in a given year. My data sample goes from 2000 to 2019. In the main sample I remove GSE loans, i.e. the mortgages subsidized by the Federal Housing Authority, the U.S. Department of Veterans Affairs, or other government programs. For the bank institutions in this dataset, I use the RSSD identifier to merge their home mortgage loan data with the Call Reports.[17]

County data. I collect data on county population, employment, and median household income from the U.S. Bureau of Labor Statistics, Bereau of Economic Analysis (BEA) and the Census Bureau. I match the data to other datasets using the county fips code as the identifier. Information on local business activities such as two-digit-industry level employment and number of establishments is provided by the County Business Patterns.

Monetary policy data. The quarterly data of effective Fed funds rates and treasury yield rates are obtained from the Federal Reserve Economic Data (FRED). For identification issue, I also adopt the series for policy news shocks from Nakamura and Steinsson (2018) and information shocks from Jarociński and Karadi (2020) as instruments for the Fed funds rate.[18] The sample is all regularly scheduled FOMC meetings from January 2000 to December 2019, excluding the peak of the financial crisis from July 2008 to June 2009. Following Romer and

[17]The RSSD identifier is provided by Robert Avery from the Federal Housing Finance Agency (available at https://sites.google.com/site/neilbhutta/data).

[18]I use the updated series by Acosta and Saia over January 2000 to December 2019. The shock series is the first principal component across surprise changes of five futures contracts around scheduled policy announcements: the one with respect to the Fed funds rate immediately following a meeting by the Federal Open Market Committee (FOMC), the expected federal funds rate immediately following the next FOMC meeting, and expected three-month eurodollar interest rates at horizons of two, three, and four quarters.

Romer (2004), I convert the shocks data to quarterly frequency by summing up the shocks within the same quarter.

1.4.2 Variable definition

This section presents the definition of main variables for the empirical analysis. Table 1.2 in Appendix 1.C lists the additional variables as well as their summary statistics.

Nominal interest rate. In my theoretical model, the nominal interest rate is the marginal cost of loans. The model assumes it is equal to (or at least influenced by) the policy rate set by the central bank. However, since bank loans have longer maturity than wholesale funding, the marginal cost of loans is not necessarily equal to the policy interest rate in the data due to term premia. To account for the term premia, I follow Wang et al. (2020) to construct a Treasury portfolio, which replicates the repricing maturity of the aggregate loan portfolio of U.S. banks as reported in the Call Reports.[19] I use the weighted average yield rate of this portfolio as banks' effective nominal interest rate. This yield rate is a better measure of nominal interest rate than the Fed funds rate, since the Fed funds rate is not informative on banks' marginal cost of capital after the 2008 global financial crisis.[20]

The top panel of Figure 1.4 plots the time series of the yield rate of the aggregate replicating portfolio. It shows that the yield rate is highly correlated with the effective

[19]Since banks can also invest deposit funds in the replicating portfolio, we can think of the yield rate of this portfolio as marginal return of deposits. The repricing and maturity data for loans are reported in the Memoranda of Schedule RC-C Part I of the Call Reports. The schedule divides a bank's loan portfolio by the remaining maturity into six categories: 3 months or less, 3 months to 12 months, 1 to 3 years, 3 to 5 years, 5 to 15 years and over 15 years. I assign to each category the yield rate of treasury that has the closest maturity, and compute the weighted average of the treasury yield rate. This procedure follows English et al. (2018). Similar calculation is adopted in Wang (2018), Begenau and Stafford (2019) and Drechsler et al. (forthcoming).

[20]Due to the unconventional monetary policy adopted by the Federal Reserve, domestic banks are flushed with excess reserves after the financial crisis, and the main participants of the Fed funds market are government sponsored enterprises and foreign bank organizations. The effective Fed funds rate falls below IOER due to regulatory arbitrage (e.g. Duffie and Krishnamurthy, 2016; Armenter and Lester, 2017; Afonso et al., 2019).

Fed funds rate, for both the periods before and after the global financial crisis. Moreover, although banks differ in their individual loan portfolio, the yield rate of the replicating portfolio does not have a large variation across banks. In the bottom panel of Figure 1.4, I calculate the yield rate of the replicating portfolio for each individual bank based on its loan maturity structure and plot the time series of quartiles. The difference between the quartiles is stable and close to 0 over time.

Loan market concentration. In our empirical analysis, the primary proxy for the loan market concentration is the concentration of residential mortgage loans, which is measured as the standard Herfindahl index (HHI) of the home mortgage loans from the HMDA data. It is calculated by summing up the squared loan-market shares of all financial institutions that originate or purchase home mortgage loans in a given county in a given year. I assign to each bank branch in a given year the HHI of the county in which it is located, and refer to it as the Branch-HMDA-HHI. Then for each bank in a given year, I take the weighted average of Branch-HMDA-HHI across its branches, using branch mortgage loan volume as weights, and refer to it as the Bank-HMDA-HHI.

The measure of loan market concentration deserves specific discussions. First, the ideal construction of HHI in local banking markets should use the information on all types of bank loans. However, only the data of home mortage loans (HMDA) and small business loans (Community Reinvestment Act, CRA) are publicly available at the bank-county-year level. I use the mortgage loan data instead of small business loan data for two reasons. First, mortgages loans account for the most substantial part of bank loans, while the share of small business loans is small. Mankart et al. (2020) use the 2010 Call Reports data and find that mortgages account for between 62% and 72% of all bank loans, while the share of commercial & industrial loans is about 10%. Thus the level of mortgage market concentration is an important determinant of a bank's overal loan market concentration. The second reason is that the number of banks reporting CRA data is much less than that of HMDA reporting

banks. When matched with the Call Reports data, the fraction of bank-year pairs with unmissing loan observations is 13% for CRA data, and 39% for HMDA data.[21]

Second, I define banking markets in this paper to be counties, which are the primary administrative divisions for most states. Although other market definitions, such as state or metropolitan statistical area, have been used in some existing empirical research on the U.S. banking industry, many on bank market power have considered county as their measure of geographic market (e.g. Drechsler et al., 2017; Chen et al., 2017; Scharfstein and Sunderam, 2016; Aguirregabiria et al., 2019).

Deposit growth. The branch-level deposit growth is obtained from the FDIC data. Since the data are reported annually, a branch's deposit growth is equal to the log difference of its deposit volume in a year. The bank-level deposit growth is obtained from the Call Reports, which is reported at quarter frequency. A bank's deposit growth is the log difference of the bank's total domestic deposits in a quarter.

Deposit rates. The branch-level and bank-level deposit rates are measured at quarterly frequency. At branch level, for each deposit product, a branch's quarterly deposit rate is equal to the quarterly average of the branch's weekly deposit rates in Ratewatch. The bank-level deposit rates are calculated from the Call Reports data. It is equal to the annualized quarterly interest expenses on domestic deposits divided by total domestic deposits.

[21]There are some restrictions on loan reporting in both data. Under CRA, all banks with assets greater than $1 billion (before 2005, it is $250 million) are required to disclose annual tract-level data on the number and dollar volume of loans originated to businesses with gross annual revenues less than or equal to $1 million. Under HMDA reporting criteria, financial institutions required to disclose are banks, credit unions and savings associations that have at least $43 million in assets, have a branch office in a metropolitan statistical area or metropolitan division, originated at least one home purchase loan or refinancing of a home purchase loan in the preceding calendar year, and are federally insured or regulated. However, as reported in Greenstone et al. (2020), the CRA data still account for 86% of total lending in the small business loans, and the HMDA data account for at least 83% of the population lived in an MSA region during the sample period.

Bank equity. The measure of bank equity is calculated from the Call Reports data. It is equal to a bank's total equity capital divided by total assets in a quarter. This is an equivalent measure of bank leverage as documented in English et al. (2018).

1.4.3 Summary statistics

Table 1.2 in Appendix 1.C provides the summary statistics of the main variables. It reports the statistics over the full sample, as well as the subperiods before 2010 (the period before low interest rate) and after 2010 (low interest rate period). Panel A presents the summary statistics for the quarterly changes of branch deposit rates. For each deposit product, the quarterly changes of deposit rates are negative on average, which is due to the long-run decline of nominal interest rate. Moreover, the deposit rates before 2010 decrease more and are more volatile than the deposit rates after 2010. This is because the nominal interest rate is less volatile in the low-interest rate period.

Panel B reports the summary statistics of the annual deposit growth of U.S. branches. We can observe that the average growth rate is higher before 2010 than after 2010. This is because the nominal interest rate decreases more on average before 2010 than after 2010.

Panel C presents the summary statistics for the home mortgage loan HHI of counties with at least one bank branch. The average HHI is low, implying that the home mortgage loan markets are quite competitive on average. However, the standard deviation is close to the mean value, which implies a large variation of HHI across counties. These results are also reflected in Figure 1.5, which plots the map of Branch-HMDA-HHI across counties in the United States. Moreover, the average HHI of home mortgage loan markets increases from 0.08 in the first subperiod to 0.12 in the second subperiod, with a slight reduction in the standard deviation.

Panel D reports the sumamry statistics for bank characteristics. Similar to the summary statistics of branch deposit rates, the quarterly changes of bank deposit rate is negative

on average. The bank deposit rates decrease more and are more volatile before 2010 than after 2010. The bank-level average home mortgage loan HHI (Bank-HMDA-HHI) is smaller on average and less volatile than the measures at county level. The distribution of bank equity-assets ratio is concentrated and quite stable across periods. Finally, there is a large reduction in the number of banks across periods due to the waves of bank failure and mergers and acquisitions.

Panel E presents the summary statistics of the yield rate on the aggregate replicating portfolio. Consistent with the trend of Fed funds rate, the yield rate is lower and less volatile over the years after 2010 than the period before 2010.

1.5 Empirical analysis

This section presents the empirical tests of our model. The analysis starts with branch-level regressions that identify the theoretical mechanism, and then provides bank-level estimation that documents the impact of our channel on bank balance sheets.

1.5.1 Branch-level estimation

The branch-level estimation aims to verify the testable predictions (1.41) and (1.42). The detailed description of identification strategy and empirical results are presented below.

1.5.1.1 Identification assumption

The first part of our empirical analysis is to identify the causal effect of loan market competition on the passthrough to deposit rates under low interest rate. The main identification issue is that the changes in deposit rates and volumes depend on the loan and deposit opportunities simultaneously. In order to guarantee that banks are faced with similar deposit opportunities, I compare the deposit rates and deposit volume growth across branches in the

same county but belong to different parent banks.[22] This identification strategy assumes that banks can raise deposits at one branch and lend them at another to equalize the marginal returns of lending across branches. It implies that the impact of loan market competition on a bank's deposit rate is determined by the average loan market concentration of its branches. Therefore, the within-county estimation is able to control for the branches' deposit market power and identify the effect of loan market competition on the passthrough to deposit rates. The identifying assumption is empirically justified by Drechsler et al. (2017), who show that a bank's lending in a given county is not related to local deposit-market concentration. The related empirical evidence is also documented in the banking literature, which shows that banks reallocate deposit fundings to areas with high loan demand (Gilje et al., 2016). Moreover, the within-county estimation allows me to control any other local market characteristics that can affect the equilibrium deposit rates.

1.5.1.2 Preliminary analysis

The preliminary analysis investigates if the relationship between passthrough and loan market concentration is different, when the nominal interest rate is high versus low. To do this, I split the data sample into two subperiods: the quarters before 2010Q1 and the quarters after 2010Q1, and run the following regression:

$$\begin{aligned} \Delta y_{j,t} &= \alpha_j + \gamma_{b(j)} + \delta_{c(j),t} + \beta_1 1\{t < 2010\text{Q1}\} \times HHI_{b(j),t-1} \times \Delta i_t \\ &\quad + \beta_2 1\{t \geq 2010\text{Q1}\} \times HHI_{b(j),t-1} \times \Delta i_t + \beta_3 HHI_{b(j),t-1} + \gamma \cdot \mathbf{x}_{b(j),t-1} + \varepsilon_{j,t}, \end{aligned} \tag{1.43}$$

where j denotes a branch, $b(j)$ denotes the parent bank, $c(j)$ denotes the county of branch j and t is the time index (quarter). I include county-time fixed effect $\delta_{c(j),t}$ to implement the within-county estimation, and branch fixed effect α_j and bank fixed effect $\gamma_{b(j)}$ to control the

[22] This identification is similar to the within-county estimation on the impact of deposits concentration on small business lending in Drechsler et al. (2017).

unobserved time-invariant characteristics of branches and parent banks.[23] The variable $y_{j,t}$ is either the deposit rate or the log of deposits of branch i in period t, and $\Delta y_{j,t}$ is the change of $y_{j,t}$ from period $t-1$ to t. The variable Δi_t is the contemporaneous change in the Fed funds rate. The variable $HHI_{b(j),t}$ is the measure of bank-level loan market concentration, i.e. Bank-HMDA-HHI. The indicator functions $1\{t < 2010Q1\}$ and $1\{t \geq 2010Q1\}$ represent the dummies of two subperiods. The interaction terms capture the heterogeneous impact of loan market concentration on the deposit rate passthrough.[24] I use quarterly data for deposit rates from Ratewatch and annual data for deposit growth from FDIC. I focus on the sample of counties with at least two different banks for identifying β_1 and β_2. In addition to the fixed effects, the regression also includes a set of bank characteristics $\mathbf{x}_{b(j),t-1}$ to control the factors potentially correlated with $HHI_{b(j),t}$.[25]

Table 1.3 and 1.4 report the results of preliminary specification (1.43) for deposit rates and deposit growth, respectively. The deposit rates include savings deposits ($25K money market account) and time deposits ($10K 3-month, 6-month and 12-month CDs). The deposit growth is the annual growth rate of a branch's aggregate deposits. The results of Table 1.3 show that in the period after 2010Q1, the passthrough of Fed funds rate to deposit rate is increasing in bank's loan market HHI. The results are significant for the three products of time deposits, and also produce the correct signs of coefficients for the savings deposits. However, in the period before 2010Q1, the passthrough is decreasing in bank's loan market HHI. These results are consistent with Proposition 1.2. Moreover, the magnitude of estimation is also considerable. The estimated coefficients imply that, for example, when

[23]Some branches change their ownership structure due to merger and acquisitions during the sample period. I introduce bank fixed effect to control the potential time-invariant effect of this ownership change.

[24]The results are robust if the term $\beta_3 HHI_{b(j),t-1}$ is replaced with $\beta_3 1\{t < 2010Q1\} \times HHI_{b(j),t-1} + \beta_4 1\{t \geq 2010Q1\} \times HHI_{b(j),t-1}$ to capture the heterogeneous direct effect of loan market concentration on the changes of deposit rates over two subperiods.

[25]The controls include one-period lag of bank-level average deposits HHI, its interaction with Δi_t, one-period lag of the deposit rate, and one-period lags of bank characteristics: bank size (log of assets), loan-assets ratio, share of non-performing loans, repricing maturity, core deposit share, equity-assets ratio and noninterest net income to assets ratio.

the Fed funds rate rises by 100 bps after 2010Q1, banks in high-concentration loan markets (HHI=1) raise deposit rates by 79 bps more than banks in low-concentration loan markets (HHI=0). Note that the results are robust for various specifications on fixed effects and bank controls. For the regressions on deposit growth, Table 1.4 confirm that the deposit growth increases (decreases) in loan market HHI in the period after (before) 2010Q1, when the Fed funds rate increases. This means that during the period of low interest rate, banks in high-concentration loan markets experience smaller deposit outflows than banks in low-concentration loan markets.

Equation (1.43) provides an intuitive split of the data sample to investigate the heterogeneous impact of loan market competition on the passthrough to deposit rates. However the specification is subject to several disadvantages. First, the heterogeneous impact could be driven by other factors that took place simultaneously with low interest rate. Due to the global financial crisis, there could be structural changes in banking sector that arise from changes in the bank regulations. Second, the specification cannot provide an estimate of the threshold value of nominal interest rate, below which the deposit rate passthrough increases in loan market concentration. Thus in the following section, I extend the preliminary regression to the baseline regression that explicitly takes into account the effect of nominal interest rate, as well as tesing the role of bank equity.

1.5.1.3 Baseline estimation

Our baseline regression, which is designed for testable prediction (1.41), takes on the following specification:

$$\begin{aligned}\Delta y_{j,t} &= \alpha_j + \gamma_{b(j)} + \delta_{c(j),t} + \beta_1 HHI_{b(j),t-1} \times \Delta i_t + \beta_2 HHI_{b(j),t-1} \times i_{t-1} \times \Delta i_t \\ &+\beta_3 HHI_{b(j),t-1} \times E_{b(j),t-1} \times \Delta i_t + \beta_4 E_{b(j),t-1} \times \Delta i_t + \beta_5 HHI_{b(j),t-1} \times i_{t-1} \\ &+\beta_6 E_{b(j),t-1} \times i_{t-1} + \beta_7 E_{b(j),t-1} \times HHI_{b(j),t-1} + \beta_8 HHI_{b(j),t-1} \\ &+\beta_8 E_{b(j),t-1} + \gamma \cdot \mathbf{x}_{b(j),t-1} + \varepsilon_{j,t},\end{aligned} \tag{1.44}$$

where i_t is the yield rate on the replicating treasury portfolio, and $E_{b(j),t}$ represents the bank equity-assets ratio. All the other variables are defined in the same way as in equation (1.43).

Equation (1.44) addresses the disadvantages of (1.43). First, the two-way interaction $HHI_{b(j),t-1} \times \Delta i_t$ and the three-way interaction $HHI_{b(j),t-1} \times i_{t-1} \times \Delta i_t$ capture that the effect of loan market concentration on passthrough depends on the level of nominal interest rate. The coefficient β_1 captures the sensitivity of passthrough to loan market concentration when the interet rate is zero. The coefficient β_2 measures the the change of the sensitivity when the nominal interest rate increases by 100 bps. Moreover, the term $HHI_{b(j),t-1} \times E_{b(j),t-1} \times \Delta i_t$ represents that the effect of loan market concentration on passthrough also depends on the level of bank equity. As predicted by the model, we expect $\beta_1 > 0$ and $\beta_2, \beta_3 < 0$. This means when the nominal interest rate is sufficiently low (high), the deposit rate passthrough is increasing (decreasing) in banks' loan market concentration. The threshold value of i_{t-1} is equal to $-\frac{\beta_1}{\beta_2} - \frac{\beta_3}{\beta_2} E_{b(j),t-1}$, which is decreasing in bank equity. I also add all the other two-way interactions among the main regressors, i.e., $HHI_{j(i),t-1}$, Δi_t, i_{t-1}, and $E_{j(i),t-1}$, as long as they are not absorbed by fixed effects. The standard errors are clustered at the county level.

The estimation in first differences performed in the equation (1.44) is preferable to estima-

tion in levels in our empirical analysis. The main reason is that we focus on the passthrough of policy rates to deposit rates, which is the sensitivity of deposit rates to policy rate shocks.[26] Therefore, regression in first differences is the direct counterpart of Proposition 1.2. Thus I adopt the first-difference estimation in line with the previous literature.

The estiation results of branch deposit rates are reported in Table 1.5. Column (1), (4) (7) and (10) only include the two-way interaction $HHI_{b(j),t-1} \times \Delta i_t$ to estimate the average effect of loan market concentration on the passthrough to deposit rates. All the four columns report insignificant β_1, implying that an insignificant average impact. Column (2), (5), (8) and (11) reports the baseline specification, which include the interaction terms $HHI_{b(j),t-1} \times \Delta i_t$ and $HHI_{b(j),t-1} \times i_{t-1} \times \Delta i_t$. Although the estimates in column (2) are insignificant, column (5), (8) and (11) confirm that the passthrough of Fed funds rate to time deposit rates is increasing in the loan market concentration, when the nominal interest rate is low. The differential effect vanishes as the nominal interest rate is higher. For example, when the nominal interest rate is zero and the Fed funds rate increases by 100 bps, banks in high-concentration loan markets (HHI=1) raise the deposit rates of 12-month CD accounts by 201 bps more than banks in low-concentration loan markets (HHI=0). The differential effect becomes zero when the nominal interest rate increases to 2.95%. Also notice that the absolute value of β_1's in colum (5), (8) and (11) is larger for products with longer maturity, meaning that the differential effect at zero nominal rate is stronger for longer-maturity products. Column (3), (6), (9) and (12) present the results of the full specification. The estimation demonstrates the role of bank equity. Column (3) reports a significantly positive β_1 and a significantly negative β_3, meaning that the differential effect at zero interest rate also exists for savings deposits, and the effect is weakened by a larger bank-equity ratio. It also implies that the insignificant coefficients in column (2) are due to omitted variables.

[26] Another reason of estimating in first differences is documented in Drechsler et al. (2017), which implicitly assumes that bank retail rates adjust concemporaneously to changes in the Fed funds rate. This is preferable to estimation in levels from an identification standpoint because it controls for other factors that might vary with monetary policy over longer periods of time or with a lag.

Moreover, column (6) and (12) also report an significant and negative β_3, which demonstrates the role of bank equity in generating the differential effect of loan market concentration on deposit rate passthrough.

Next I present the results that verify testable prediction (1.42). Specifically, I replace the county×time fixed effects with county fixed effects in equation (1.44), and add Δi_t, i_{t-1} and $\Delta i_t \times i_{t-1}$ to capture the aggregate effect of low interest rate on deposit rate passthrough. I also includes a set of controls for robustness of estimation.[27] Table 1.12 reports the estimation results for branch deposit rates. The coefficient of $\Delta i_t \times i_{t-1}$ is positive, which implies that for banks in perfectly competitive loan market, a lower interest rate weakens the deposit rate passthrough. This is consistent with the empirical result of Wang et al. (2020). On the other hand, the coefficient of $HHI_{j(i),t-1} \times i_{t-1} \times \Delta i_t$ is negative and significant for time deposit products. Moreover, the absolute value of the coefficient of $HHI_{j(i),t-1} \times i_{t-1} \times \Delta i_t$ is larger than that of $\Delta i_t \times i_{t-1}$. This implies that when a bank's loan market concentration is sufficiently large, a lower nominal interest rate actually improves the deposit rate passthrough. The estimated coefficients reveal that a lower nominal rate can improve the passthrough of all the time deposit products, if the bank's loan market HHI is above 0.21. These estimation results show that the impact of low interest rate on passthrough is heterogeneous and depends on banks' loan market power.

Table 1.13 presents within-county estimates for deposit volume growth. Column (1) to (3) verify testable prediction (1.41): when the nominal interest rate is zero, an increase in the Fed funds rate leads to smaller outflows for the banks in more concentrated loan markets. The differential effect shrinks at a higher nominal interest rate. Column (3) shows that the threshold nominal interest rate is decreasing in bank equity-assets ratio. For banks with average equity-assets ratio, a 100 bps Fed funds rate increase from zero generates 389 bps

[27]The set of controls includes 4-quarter lags of retail and treasury rates, 4-quarter lags of unemployment and real GDP growth, one-year lag of Branch-Dep-HHI, county share of population aged 65 or older, log of county-level population, log of county-level median household income and county share of population with a college degree. I also add linear and quadratic time trends, and the interactions of these controls with the monetary shock.

less deposit outflows if the bank locates at high-concentration loan markets (HHI=1) than at low-concentration loan markets (HHI=0). Moreover, column (8) reports the results of testable prediction (1.42). I find that a lower interest rate increases the sensitivity of deposit growth to Fed funds rate for banks in high-concentration loan markets, but decreases the sensitivity for banks in low-concentration loan markets.

1.5.1.4 Other determinants of the channel

The theoretical model and empirical results show that the threshold value of nominal interest rate, below which the deposit rate passthrough increases in loan market concentration, is decreasing in bank equity-assets ratio. In this section I investigate if other bank characteristics also affect the threshold value of nominal interest rate.[28] Specifically, I run the following regression:

$$\begin{aligned}\Delta y_{j,t} &= \alpha_j + \gamma_{b(j)} + \delta_{c(j),t} + \beta_1 HHI_{b(j),t-1} \times \Delta i_t + \beta_2 HHI_{b(j),t-1} \times i_{t-1} \times \Delta i_t \quad (1.45)\\ &+\beta_3 HHI_{b(j),t-1} \times i_{t-1} + \beta_4 HHI_{b(j),t-1} + \theta_1 \cdot HHI_{b(j),t-1} \times \mathbf{x}_{b(j),t-1} \times \Delta i_t\\ &+\theta_2 \cdot \mathbf{x}_{b(j),t-1} \times \Delta i_t + \theta_3 \cdot \mathbf{x}_{b(j),t-1} \times i_{t-1} + \theta_4 \cdot \mathbf{x}_{b(j),t-1} \times HHI_{b(j),t-1}\\ &+\gamma \cdot \mathbf{x}_{b(j),t-1} + \varepsilon_{j,t},\end{aligned}$$

where I replace $E_{b(j),t-1}$ with a full set of bank characteristics $\mathbf{x}_{b(j),t-1}$. The set of bank characteristics include equity-assets ratio, log of bank assets, loan-assets ratio, core deposit share in total liabilities, maturity gap, nonperforming share of loans in total loans, the share of other assets in total interest-earning assets, the share of other liabilities in total liabilities. The computation of maturity gap follows English et al. (2018). The variables "other assets" and "other liabilities" represent the assets and liabilities with no repricing

[28]The potential determinants reflect other financial frictions that link banks' loan pricing strategy to deposits pricing. Such frictions include borrowing cost in wholesale funding market due to asymmetric information, regulatory requirement on reserves, capital and liquid assets. See Wang et al. (2020) for a quantitative evaluation on the role of each friction in monetary transmission.

or maturity information. Including these three variables aims to investigate the impact of bank's maturity structure on the threshold value of nominal interest rate.

Table 1.15 reports the estimated vector of coefficients θ_1 for branch deposit spreads. Consistent with the baseline estimation, the coefficients of three-way interaction $HHI_{j(i),t-1} \times E_{j(i),t-1} \times \Delta i_t$ are still significant for money market accounts, 3-month CDs and 12-month CDs in this table. This confirms that our theoretical mechanism through capital constraint is significant for linking loan market concentration and deposit pricing, and bank equity is one of the most important determinants for the threshold value of nominal interest rate. The effects of other bank characteristics are as follows. First, the loan market concentration has a significantly more pronounced positive impact on larger banks' deposit rate passthrough, as evidenced by the large positive coefficient of bank size. This implies that the threshold value of nominal interest rate is increasing in bank assets. This is possibly due to stronger regulatory capital constraint on larger banks. Second, as indicated by the positive coefficients of nonperforming loan share, loan market concentration also improves the deposit passthrough more on banks with a larger share of nonperforming loans. A possible reason is that higher nonperforming loan share implies higher risk of bank assets, which induces a more strict capital constraint or higher cost of external financing. Thus the threshold value of nominal interest rate is also increasing in nonperforming loan share.

1.5.1.5 Robustness checks

This section presents the robustness checks of the empirical results. First, the results are robust for alternative fixed effects and controls. Table 1.6 and 1.7 report the robustness checks for deposit rates, and column (4)-(7) in Table 1.13 report the robustness checks for deposit growth. Second, the results are similar if the measure of nominal interest rate is the Fed funds rate or the 1-year treasury yield rate. Table 1.8 and 1.9 report the estimation results for deposit rates, and column (1)-(4) in Table 1.14 report the results for deposit growth. Third, restricting the data sample to the pre-financial crisis period (until the end of

2008Q2) produces consistent signs of coefficients, but slightly reduces the significance level. This is because the nominal interest rate was not extremely low. The results are reported in Table 1.10 for deposit rates, and column (5) and (6) in Table 1.14 for deposit growth. Fourth, the robustness tests run the original regressions of (1.44) for banks with the largest 25% banks, which are sorted by the inflation-adjusted annual average assets, and obtain consistent results. The regression results are reported in Table 1.11 for deposit rates, and column (7) and (8) in Table 1.14 for deposit growth.All these results are availalbe upon request.

1.5.2 Bank-level estimation

To deepen the understanding of the economic consequences of low interest rate and bank market power on the passthrough, I now turn to banks' income and balance sheet variables. In particular, I investigate whether the results of branch-level regressions also hold at the bank level, and how the impact of loan market competition on deposit passthrough affects the associated dynamics of bank balance sheets.[29] This provides a more detailed picture of how banks with different loan market power respond differently to interest rate changes in the low-interest environment. The empirical analysis uses the Call Report data over 1997Q1 to 2019Q4.

The analysis studies the impact of loan market concentration on the sensitivity of bank balance sheet components to monetary policy, under different levels of nominal interest rates.

[29]I focus on the dynamic responses because the banks' income and balance sheet reflect income and expense flows accruing from past assets and liabilities. The impacts of interest rate shocks can be reflected on the data with lags.

Formally, I estimate the following bank-quarter regression:

$$\begin{aligned} y_{b,t+h} - y_{b,t-1} &= \alpha_{b,h} + \eta_{t,h} + \beta_{1,h} HHI_{b,t-1} \times \Delta i_t + \beta_{2,h} HHI_{b,t-1} \times i_{t-1} \times \Delta i_t \\ &+ \beta_{3,h} HHI_{b,t-1} \times E_{b,t-1} \times \Delta i_t + \beta_{4,h} E_{b,t-1} \times \Delta i_t \\ &+ \beta_{5,h} HHI_{b,t-1} \times i_{t-1} + \beta_{6,h} E_{b,t-1} \times HHI_{b,t-1} \\ &+ \beta_{7,h} HHI_{b,t-1} + \beta_{8,h} E_{b,t-1} + \gamma_h \cdot \mathbf{x}_{b,t-1} + \varepsilon_{b,t,h}, \end{aligned} \tag{1.46}$$

where $h = 0, 1, 2, ..., 8$, b represents a bank and t is a quarter. The dependent variable $y_{b,t+h}$ is an accounting measure of a bank balance sheet variable in quarter $t + h$. The bank concentration $HHI_{b,t}$ is the bank-level loan market HHI, i.e. Bank-HMDA-HHI. The interest rate changes Δi_t is the shocks to Fed funds rate. To control the potential endogeneity between the Fed funds rate changes and the unobserved factors at the bank level, I use two sequences of monetary policy shocks: the policy news shocks of Nakamura and Steinsson (2018) and the information shocks of Jarociński and Karadi (2020). The monetary policy shocks are normalized to generate +100 bps change in the Fed funds rate. I present the results using each sequence and show the robustness. The nominal interest rate i_t and bank equity $E_{b,t}$ are defined in the same way as before. I control the horizon-specific bank fixed effect $\alpha_{j,h}$ and time fixed effect $\eta_{t,h}$, as well as bank-level controls $\mathbf{x}_{b,t-1}$.[30] The inclusion of bank controls is intended to absorb non-monetary policy drivers of bank balance sheet. I cluster standard errors at bank level. The estimation follows the local projection method of Jordà (2005). I plot the sequence of estimated coefficients $\left\{\hat{\beta}_{1,h}, \hat{\beta}_{2,h}, \hat{\beta}_{3,h}\right\}$, $h = 0, 1, ..., 8$, which traces out the cumulative response of bank-level variables to a policy-induced change in the Fed funds rate as a function of bank loan market concentration, level of nominal interest rate and bank equity.

[30]The bank controls include the same set of balance sheet variables as in the branch regressions. We also include two lags of $y_{b,t}$ to control the seasonality of dependent variables. The full results are available upon request.

Results on deposit rates. Figure 1.6 plots the estimated responses of average bank deposit rates using the policy news shocks of Nakamura and Steinsson (2018). The bank deposit rate is equal to the annualized quarterly interest expenses on domestic deposits divided by total domestic deposits. Panel (a)-(c) depict the sequences of $\left\{\hat{\beta}_{1,h}\right\}$, $\left\{\hat{\beta}_{2,h}\right\}$ and $\left\{\hat{\beta}_{3,h}\right\}$ respectively. Consistent with the branch-level estimation, the impact of loan market concentration and nominal interest rate is significant persistent on the passthrough to deposit rate at bank level. As shown in panel (a), when there is a +100 bps change in the nominal interest rate from zero-lower bound, banks which has zero equity and operates in a high-concentration loan market (HHI=1) increase their deposit rates by an average of 203 bps more than banks which has zero equity and locates in a low-concentration loan market (HHI=0), during the subsequent four quarters. The estimate is similar in magnitude to the branch-level estimation of Table 1.5. Panel (b) shows that the differential response vanishes by an average of 29.6 bps if the level of nominal interest rate increases by 100 bps, which is slightly smaller than the magnitude estimated in the branch regression. Panel (c) shows that the differential response vanishes by about 12.9 bps if the bank equity-to-assets ratio increases by 1%. The estimates imply that the threshold level of nominal interest rate, below which banks with a higher loan market concentration have more efficient passthrough, is about 1.89%. The number is lower than the estimate values on time deposits in the branch regressions, which is because the bank average deposit rates also include savings and demand deposits. Moreover, when the nominal interest rate is equal to 3.93%, i.e. its pre-crisis average (over 1997Q1 to 2007Q4), then the banks with average equity-assets ratio and high loan market concentration (HHI=1) raise deposit rates by an average of 60.4 bps less than those with average equity-assets ratio and low loan market concentration (HHI=0). All these numbers imply that the magnitudes of the responses under both normal and low interest rates are economically meaningful compared to the standard deposits channel.[31] All

[31]For a more reasonable comparison, the banks with average equity-assets ratio raise deposit rates by 6.1 bps more and 1.8 bps less if their loan market HHIs are higher by one standard deviation, when the interest rate increases by 100 bps from zero and the pre-crisis average, respectively. The corresponding number

these estimates are significant at least 5% level for a horizon of four quarters, and gradually shrink to zero over the following quarters. This implies that our channel on the short-run passthrough lasts about four quarters, which is consistent with the fact that the maturities of bank deposit products are mostly below 12 months. Moreover, the results are robust and significant if I use the information shocks from Jarociński and Karadi (2020) as monetary policy shocks. As show in Figure 1.10, the estimated coefficients of $\left\{\hat{\beta}_{1,h}, \hat{\beta}_{2,h}, \hat{\beta}_{3,h}\right\}$ are of the correct signs and significant for the first two quarters.

Results on balance sheet growth and structure. Now I turn to estimate the impact of our channel on the dynamics of bank balance sheet variables. The analysis focuses on estimating the responses of the growth of bank deposits, loans, securities and assets, as well as the loan-security ratio and the core deposit share in total liabilities. For deposits, loans, securities and assets, which is denoted as $Y_{j,t}$, I define the dependent variable $y_{j,t+h}$ at horizon h as the symmetric growth rate of $Y_{j,t}$ between $t-1$ and $t+h$, i.e. $y_{j,t+h} = \frac{Y_{j,t+h}-Y_{j,t-1}}{0.5\left(Y_{j,t+h},Y_{j,t-1}\right)}$.[32] The symmetric growth rate is able to accomodate changes in bank balance sheet variables from a starting level of zero, and bound the changes between -2 and 2 for all impulse response horizons and therefore avoids the possibility of extreme outliers. All the other specifications stay the same with (1.46).

Figure 1.7 reports the estimation results using the monetary policy shocks of Nakamura and Steinsson (2018). It shows that the our channel has significant impact on the dynamics of bank balance sheets. As plotted in panel (a), the effects of loan market concentration and bank equity are both significant and of the correct signs over the second and third quater after the interest rate shock, and the effect of the level of nominal interest rate is negative

estimtated in Drechsler et al. (2017) is by 1.1 bps less if the banks' deposit HHIs are one standard deviation higher.

[32]The symmetric growth rate is the second-order approximation of the log-difference for growth rates around zero and has been used in a variety of contexts such as establishment-level employment growth rates (e.g. Decker et al. (2014)) but also credit growth rates (e.g., Gomez et al., 2021 and Greenwald et al., 2020).

and significant over the quarter 0 and 1. Panel (b) to (d) show that banks with higher loan market concentration absorb this expansion by increasing their loans and securities more, when the nominal interest rate is zero. For the growth of bank loans, the level of nominal interest rate does not have significant impact on the passthrough. However, the level of nominal interest rate has a significant impact on the passthrough to assets and security holdings growth in the contemporaneous quarter when the interest rate shocks take place. Moreover, the results with the shocks of Jarociński and Karadi (2020) show that the level of nominal interest rate does have significant impact on the growth of bank balance sheets. As reported in Figure 1.11, this effect is significant with at least 5% level for the growth of deposits, assets and securities over two quarters after the interest rate shock, and for the loan growth over the third and fourth quarter after the interest rate shock. One reason is that the shocks of Nakamura and Steinsson (2018) rely more on the yield rate curves with shorter maturity than the shocks of Jarociński and Karadi (2020), thus does not fully reflect the shocks to the cost of bank capital.

Finally, the composition of bank balance sheet is also impacted by our passthrough channel. Figure 1.7 and 1.11 report the responses of a bank's loan-assets ratio and the share of core deposits in liability. In response to an increase in the Fed funds rate from zero, banks with higher loan market concentration experience larger increase in both loan-assets ratio and core deposit share. The differential response of core deposit share is due to the differential effect of loan market competition on deposit passthrough. Banks with higher loan market concentration are able to raise more deposits, which gives rise to a higher share of core deposits. Since the cost of raising core deposits is lower than other funds, these banks are able to supply loans at lower average and marginal costs. Then they obtain an advantage of issuing new loans and experience an increase in the share of loans in assets.

Results on profitability. A major concern of low interest rate is that it narrows bank profitability and thereby reduces bank lending through the leverage constraints.Thus I inves-

tigate the impact of the reversed deposits channel on the dynamics of the measures of bank profitability. Specifically, I run the regression of (1.46) on the equity-assets ratio and net interest margin. The net interest margin is equal to the annualized quarterly interest income on assets minus quarterly interest expense on liabilities, and then divided by bank assets. The results are plotted in Figure 1.9 and 1.13. In both figures, the level of nominal interest rate has negative and persistent effect on the dynamics of bank equity-assets ratio through loan market competition. For example, in Figure 1.9, at zero nominal interest rate, there is no significant heterogeneity in the response of equity-assets ratio to interest rate shocks for banks with different loan market concentration. However, when the nominal interest rate is equal to 1%, the equity-assets ratio of banks with high loan market concentration (HHI=1) reduces more than banks with low loan market concentration (HHI=0) to positive interest rate shocks, and this additional reduction persists and accumulates to 0.03% in two years after the shock. The magnitude is considerable since the standard deviation of equity-assets ratio is 0.07%.

Moreover, the differential effect of loan market concentration and nominal interest rate also exists for the net interest margin. As shown in panel (b) of both figures, banks with higher loan market concentration experience larger change of net interest margin in response to interest rate shocks when the nominal rate is zero. The differential response vanishes as the nominal rate gets higher. This is consistent with the responses of loan-assets ratio and equity-assets ratio: since the loan-assets ratio increases more for the banks with higher loan market concentration, these banks gain more interest income on new loans, which offsets the loss on deposits. This increases the banks' net interest margin and then the bank equity. When the interest rate shock takes place at a high level of nominal interest rate, banks are not able to gain more on loan interest income, thus the contribution of loan market concentration on profitability is weakenend. This gives rise to a decreasing response of bank equity and net interest margin over the level of nominal interest rates.

1.5.3 State-dependent exposure to monetary policy

The above analysis shows that the passthrough of monetary policy to deposit rates depends on the level of nominal interest rate. The theoretical model shows that this dependence operate through loan market competition. Yet the dependence could also derive from other sources. Moreover, the bank-level regressions suggest that this dependence is passed through to the responses of bank balance sheets upon interest rate shocks. In this section, I propose a state-dependent measure of passthrough, which comprehensively evaluates the impact of nominal interest rate on monetary policy transmission through bank balance sheets.

The measure builds on the deposit spread beta in Drechsler et al. (2017). Specifically, I assume the passthrough of monetary policy to deposit rate is a linear function in the nominal interest rate. This is done by running the following time series regression for each bank in the Call Reports data:

$$\Delta\text{Deposit Rate}_{bt} = \alpha_b + \sum_{\tau=0}^{3} \beta_{0,b,\tau} \Delta i_{t-\tau} + \sum_{\tau=0}^{3} \beta_{1,b,\tau} i_{t-\tau-1} \times \Delta i_{t-\tau} + \varepsilon_{bt}, \tag{1.47}$$

where b denotes a bank, t denotes quarter, $\Delta\text{Deposit Rate}_{bt}$ is the change in the deposit rate of bank b from period $t-1$ to t, i_t is the Fed funds rate in period t, Δi_t is the change in the Fed funds rate from period $t-1$ to period t.[33] Similar to Equation (1.1), I control 3 lags of interest rate shocks to account for the cumulative effects over a full year. However, this regression equation differs from the former by adding the interaction terms $\sum_{\tau=0}^{3} \beta_{1,b,\tau} i_{t-\tau-1} \times \Delta i_{t-\tau}$, which takes into account the dependence on the nominal rate. Our estimate of the passthrough to deposit rates consists of two betas: $\beta_{0,b} = \sum_{\tau=0}^{3} \beta_{0,b,\tau}$ and $\beta_{1,b} = \sum_{\tau=0}^{3} \beta_{1,b,\tau}$. The first beta $\beta_{0,b}$ measures the passthrough when the nominal rate is zero, thus I call it "zero beta". The second beta $\beta_{1,b}$ measures the change in the passthrough

[33] Alternatively, one can replace the Fed funds rate changes with the monetary policy shocks, or replace the lag of Fed funds rate with the lag of the yield rate of replicating portfolio. All the specifications produce similar results.

when the nominal rate increases by 100 bps, thus I call it "slope beta". Since our dependent variable is the changes in deposit rates, a positive slope beta means the passthrough is lower at a lower interest rate, and a negative slope beta means the passthrough is higher at a lower nominal rate. For the robustness of estimation, I focus on the banks which have at least 60 quarters of data over 1997Q1 to 2019Q4, excluding the periods of global financial crisis (2008Q3 to 2009Q2), and winsorize the estimated betas at the 10% level to remove the impact of outliers.

Our estimates suggest that the deposit rate passthrough at zero nominal rate is substantially low and the passthrough is quite sensitive to the nominal interest rate. For all of the banks included in the estimation, the average values of zero beta and slope beta are 0.259 and 0.030, respectively. That is, on average banks raise deposit rates by 25.9 bps per 100 bps increase in the Fed funds rate, when the initial Fed funds rate is zero. This amount increases to 37.9 bps per 100 bps in the Fed funds rate if the initial Fed funds rate increases to 4%. This implies that on average, the passthrough efficiency is lower at lower interest rate. However, the betas also differ substantially in the cross section. The standard deviations of zero beta and slope beta are 0.140 and 0.038, respectively. The fraction of banks with a negative slope beta is 24.6% (1203 out of 4885). This means that low interest rate improves the passthrough efficiency to deposit rates for a quarter of banks in our sample. For comparison, the 10th and 90th percentiles of the slope beta distribution are -0.032 and 0.090. This implies a large cross-sectional heterogeneity in the change of passthrough: when the initial Fed funds rate reduces from 4% to 0, the banks at the 10th percentile are able to raise deposit rates by 36 bps more per 100 bps increase in the Fed funds rate, while those at the 90th percentile are able to raise deposit rates by 12.8 bps less. A more striking result is that the size-weighted average slope beta, with the size equal to average bank assets, is about 0. This implies that the aggregate effect of nominal interest rate on passthrough is actually zero, if we take into account the heterogeneity of bank passthrough.

Moreover, I also relate the deposit rate slope beta to deposit zero beta. Panel (a) of

Figure 1.14 reports the bin scatter plots, which sort banks into 100 bins by their zero betas and plot the average slope beta within each bin. The slope between two betas is significantly negative. This is consistent with our channel of loan market competition, which argues that banks with lower passthrough at high nominal rate is expected to have a higher passthrough at zero nominal rate.

Cross-sectional effects on bank balance sheets. Next I show that the zero beta and slope beta of deposit rates are related to the sensitivity of bank balance sheets to monetary policy. I measure the sensitivity by re-running regression (1.47) with the symmetric growth rates of deposits, assets, securities, and loans as dependent variables. The corresponding betas are called flow zero betas and flow slope betas. I present the relationship by first showing the bin scatter plots of flow zero beta vs deposit rate zero beta and flow slope beta vs deposit rate slope beta, respectively, for the growth of deposits and loans. The slope of this relationship measures the impact of increased sensitivity of deposit rates to policy rates on the various components of bank balance sheets. The results are reported in Figure 1.15. All the panels show a strong positive relationship between the deposit rate betas and the flow betas. In particular, the effect of nominal interest rate on the exposure of bank balance sheet components through deposit rate passthrough is large: when the initial level of Fed funds rate reduces from 4% to 0, the banks at the 10th percentile of the deposit rate slope beta distribution are predicted to have a 114 bps less outflow of deposits for every 100 bps increase in the Fed funds rate; however, banks at the 90th percentile are predicted to have a 320 bps more outflow of deposits the same amount increase in the Fed funds rate. The effects are similar for total assets, securities, and loans.[34] Panel A and B of Table 1.16 report the formal estimates from cross-sectional regressions of flow betas on deposit rate betas for all the banks in our sample. The estimates are significant with large magnitudes.

[34]The corresponding numbers for total assets are 88 bps less at 10th percentile and 248 bps more at 90th percentile; for securities, the numbers are 160 bps less at 10th percentile and 450 bps more at 90th percentile; for loans, the numbers are 60 bps less at 10th percentile and 169 bps more at 90th percentile.

The estimates of zero beta can be interpreted as the semi-elasticities of bank balance sheet components to deposit rates at zero nominal rate, while the estimates of slope beta can be interpreted as the sensitivity of semi-elasticities to the nominal interest rate. These results show that the impact of nominal rate on deposit rate passthrough strongly influences the sensitivity of bank balance sheets to monetary policy.

Aggregate effects. I use the estimation on large banks to calculate the aggregate impact of nominal interest rate on the deposit rate passthrough over the cross section of banks, and the corresponding impact on bank balance sheets. Our analysis focus on the largest 5% banks by assets. The summary statistics of zero beta and slope beta for deposit rates are similar to the full sample. For these banks, the averages of zero beta and slope beta of deposit rates are 0.334 and 0.022, respectively. Their standard deviations are 0.198 and 0.053, and the fraction of large banks with a negative slope beta is 32.4% (79 out of 244). This implies that even for the large banks, a lower interest rate improves the deposit rate passthrough for a significant fraction of banks, but weakens the group on average. Panel (b) of Figure 1.14 show that the two betas are still negatively correlated for the large banks. Moreover, Figure 1.16 report the bin scatter plots between flow betas and deposit spread betas for the growth of deposits and assets of large banks. All the panels confirm a positive relationship. Panel C and D of Table 1.16 report the formal estimates of this positive relationship from cross-sectional regressions of flow betas on deposit rate betas for large banks. The estimates are all significant and have overally larger magnitudes than the estimates for full sample. Thus the passthrough to deposit rates and its dependence on the level of nominal interest rate have stronger impacts on the exposure of bank balance sheets to monetary policy, which implies a strong aggregate effect since the large banks represent most of assets, deposits and loans in the U.S. banking sector.

1.6 Conclusion

This paper documents that the level of nominal interest rate affects the passthrough of monetary policy rates to deposit rates heterogeneously across banks: with a lower nominal interest rate, the deposit rate passthrough is higher (lower) if a bank starts with a low (high) passthrough. I argue that this relationship is due to banks' loan market power and capital constraints. With market power on loans, a bank's loan profit is decreasing in nominal interest rate, and its total profit is a U-shape function in nominal interest rate. The capital constraint says a bank's deposit liabilities cannot exceed its total profits. Therefore, when the nominal interest rate is low, for banks in concentrated loan markets, a lower interest rate induces larg increase in bank profits. This allows banks to take more deposits at lower deposit spreads, thereby improves the passthrough. However, for banks in competitive loan markets, a lower interest rate still depletes bank profits, which weakens the passthrough.

I test the empirical evidence of this theoretical channel using branch-level data of U.S. banks. I control for the impacts of banks' deposit market characteristics by comparing branches of different banks located in the same local banking market. I find that when the nominal interest rate is sufficiently low, branches of banks located in more concentrated loan markets raise their deposit rates by more, and experience less deposit outflows in response to increases in policy interest rates.

Since deposits are the main source of stable funding for banks, I show that this channel also affects the response of bank balance sheet components to interest rate shocks under low interest rate. Specifically, when the nominal rate is low enough, banks in more concentrated loan markets contract their balance sheet components, such as assets, securities and lendings, by less in response to policy rate increases. Moreover, these banks also experience relative increase in profitability measured by net interest margin.

Finally I extend this channel to construct a general measure of monetary policy passthrough to deposit rates, taking into account the dependence on the level of interest rate. My

estimates suggest that nominal interest rate impacts banks' passthrough heterogeneously. Specifically, a lower interest rate improves the passthrough efficiency to deposit rates for a signficant share of banks, while weakens the passthrough for the others. Further estimation shows that the dependence of passthrough efficiency on nominal interest rates can account for the effects of nominal interest rates on the monetary policy transmission through bank balance sheets.

1.A Appendix: Proofs and derivations

1.A.1 Derivations of the model

1.A.1.1 Deposit demand block

Given banks' deposit spreads $\left\{s_j^d\right\}_{j=1}^N$, the average deposit spread is defined as

$$s^d \equiv \min_{\{D_j\}_{j=1}^N} \frac{1}{N}\sum_{j=1}^N s_j^d D_j \text{ s.t. } \left(\frac{1}{N}\sum_{j=1}^N D_j^{\frac{\sigma_d-1}{\sigma_d}}\right)^{\frac{\sigma_d}{\sigma_d-1}} = 1.$$

To solve this problem, we write the following Lagrangian

$$\mathcal{L} = \frac{1}{N}\sum_{j=1}^N s_j^d D_j + \lambda\left[1 - \left(\frac{1}{N}\sum_{j=1}^N D_j^{\frac{\sigma_d-1}{\sigma_d}}\right)^{\frac{\sigma_d}{\sigma_d-1}}\right]$$

The first-order condition with respect to D_j is

$$\frac{\partial\mathcal{L}}{\partial D_j} = \frac{1}{N}s_j^d - \lambda\frac{1}{N}\left(\frac{1}{D_j}\right)^{\frac{1}{\sigma_d}} = 0 \Rightarrow D_j = \left(\frac{\lambda}{s_j^d}\right)^{\sigma_d},$$

where I apply $\left(\frac{1}{N}\sum_{j=1}^N D_j^{\frac{\sigma_d-1}{\sigma_d}}\right)^{\frac{1}{\sigma_d-1}} = 1$. Plugging the solution to D_j into the budget constraint implies

$$\lambda = \left[\frac{1}{N}\sum_{j=1}^N \left(s_j^d\right)^{1-\sigma_d}\right]^{\frac{1}{1-\sigma_d}}.$$

Therefore, plugging the solution of D_j and λ into the objective function, we obtain

$$s^d = \frac{1}{N}\sum_{j=1}^N s_j^d D_j = \frac{1}{N}\sum_{j=1}^N s_j^d\left(\frac{\lambda}{s_j^d}\right)^{\sigma_d} = \lambda = \left[\frac{1}{N}\sum_{j=1}^N \left(s_j^d\right)^{1-\sigma_d}\right]^{\frac{1}{1-\sigma_d}},$$

Moreover, since the deposit aggregator is constant returns to scale, the individual deposit demand is proportional to aggregate deposit demand:

$$D_j = \left(\frac{s^d}{s_j^d} \right)^{\sigma_d} D.$$

For the aggreate deposit demand, it is given by

$$\max_D \theta_d \cdot \ln(D) - s^d \cdot D$$

The first-order condition implies that the solution is

$$D\left(s^d\right) = \frac{\theta_d}{s^d}.$$

Finally, the demand elasticity of individual deposits is

$$\epsilon_j^d = -\frac{\partial \ln(D_j)}{\partial \ln\left(s_j^d\right)} = \sigma_d + (1-\sigma_d)\frac{\partial \ln\left(s^d\right)}{\partial \ln\left(s_j^d\right)} = \sigma_d + \frac{1-\sigma_d}{N}\left(\frac{s_j^d}{s^d}\right)^{1-\sigma_d}.$$

Q.E.D.

1.A.1.2 Loan demand block

The derivations of the average loan rate i^l and individual loan demand have the same steps as those of average deposit spread by relabelling. The aggregate loan demand is given by the following problem:

$$\max_{c^b,h,l} \frac{\left(c^b\right)^{1-\nu} - 1}{1-\nu} - \theta_h \cdot h$$

subject to

$$c^b \leq l \text{ and } \left(1+i^l\right) l \leq h.$$

The budget constraints are both binding at optimum, thus we have $c^b = l$ ahd $h = \left(1 + i^l\right) l$. Then the first-order condition for l is

$$l^{-\nu} = \theta_h \cdot \left(1 + i^l\right),$$

which implies

$$l = \left[\theta_h \cdot \left(1 + i^l\right)\right]^{-\frac{1}{\nu}}.$$

The aggregate loan demand is

$$L\left(i^l\right) = \mu \cdot l = \mu \left[\theta_h \cdot \left(1 + i^l\right)\right]^{-\frac{1}{\nu}}.$$

Q.E.D.

1.A.2 Proof of Lemma 1.1

First we derive the optimal response function of a bank's loan rate. This is given by maximizing a bank's profit on loans:

$$\max_{i_j^l} \left(i_j^l - i\right) L_j\left(i_j^l; i_{-j}^l, i\right)$$

The first-order condition implies that

$$1 = \left(1 - \frac{1+i}{1+i_j^l}\right)\left[\sigma_l + \frac{1}{N}\left(\frac{1}{\nu} - \sigma_l\right)\left(\frac{1+i_j^l}{1+i^l}\right)^{1-\sigma_l}\right].$$

Denote $x \equiv \left(1 + i^l\right)/(1+i)$ and $x_j \equiv \left(1 + i_j^l\right)/(1+i)$. Since we focus on the equilibria with non-negative loan rates, the first-order condition implies that $x_j > 1$. Then the first-order

condition can be written as

$$\frac{x_j^{\sigma_l}}{x_j - 1} - \sigma_l x_j^{\sigma_l - 1} = \frac{1}{N}\left(\frac{1}{\nu} - \sigma_l\right) x^{\sigma_l - 1}.$$

Denote the left-hand side of the above equation as $F(x_j)$. By taking first-order derivative we can get

$$F'(x_j) = x_j^{\sigma_l - 2}\left[\sigma_l \frac{x_j}{x_j - 1} - \left(\frac{x_j}{x_j - 1}\right)^2 - \sigma_l(\sigma_l - 1)\right].$$

Note that $x_j > 1$ implies $\frac{x_j}{x_j - 1} \geq 1$. Since $\sigma_l > 1$, one can show that the function $g(y) = \sigma_l \cdot y - y^2 - \sigma_l(\sigma_l - 1) < 0$ for any $y > 1$. To show this, note that $g(y)$ is maximized at $y = \frac{\sigma_l}{2}$, and $g\left(\frac{\sigma_l}{2}\right) = \frac{3}{4}\sigma_l\left(\frac{4}{3} - \sigma_l\right)$. Moreover, we have $g(1) = -(\sigma_l - 1)^2 < 0$. If $\sigma_l < 2$, then $g(y) < 0$ for any $y \geq 1$. If $\sigma_l > 2$, then $g\left(\frac{\sigma_l}{2}\right) < 0$, which also implies $g(y) < 0$ for any $y \geq 1$. Therefore, we must have $F'(x_j) < 0$ for any $x_j > 1$, which implies that $F(x_j)$ is decreasing in x_j. Moreover, note that $F(1) = +\infty$ and $F(+\infty) = -\infty$, thus there exists a unique solution of x_j to the first-order condition. This solution is identical for any j, which implies that banks must set the same loan rate in equilibrium. By replacing i_j^l with i^l in the first-order condition, we obtain the equilibrium loan rate stated in the lemma. The corresponding profit function is calculated by plugging the equilibrium loan rate into the objective function of loan problem. **Q.E.D.**

1.A.3 Proof of Lemma 1.2

First we show that the capital constraint must be binding in a symmetric equilibrium. Suppose there is a symmetric equilibrium where the constraint is not binding, then the first-order condition for individual deposit spread is

$$D_j + s_j^d D_j' = D_j\left[1 - \epsilon_j^d\right] = 0$$

which implies that

$$s_j^d = N^{\frac{1}{1-\sigma_d}} s^d.$$

Thus all banks set the same deposit spread. In this case, the equilibrium deposit spread must be zero. Otherwise, by the definition of s^d, we have

$$s^d = \left[\frac{1}{N}\sum_{j=1}^{N}\left(s_j^d\right)^{1-\sigma_d}\right]^{\frac{1}{1-\sigma_d}} = N^{\frac{1}{1-\sigma_d}} s^d \Rightarrow s^d = 0,$$

which is a contradiction. This implies that in a symmetric equilibrium with non-binding constraint, banks set zero deposit spreads, and the individual loan demand is $\frac{\theta_d}{0} = +\infty$. However, the deposit profit is a constant θ_d due to unit elastic aggregate demand. This violates the capital constraint, thus the constraint must be binding in equilibrium.

Therefore, the equilibrium deposit spread is given by the binding capital constraint, which gives the solution described in the lemma. **Q.E.D.**

1.A.4 Proof of Proposition 1.1

Taking the first-order derivative of s^d with respect to i, we have

$$\frac{\partial s^d}{\partial i} = \frac{\psi\theta_d}{\left[\mathcal{C}_l\left(1+i\right)^{-\frac{1-\nu}{\nu}} + \theta_d + \left(1+i\right)E_0\right]^2}\left\{\frac{1-\nu}{\nu}\mathcal{C}_l\left(1+i\right)^{-\frac{1}{\nu}} - E_0\right\}.$$

This implies that $\frac{\partial s^d}{\partial i} > 0$ if and only if $\frac{1-\nu}{\nu}\mathcal{C}_l\left(1+i\right)^{-\frac{1}{\nu}} - E_0 > 0$, which is equivalent to $i < i_0\left(\mathcal{C}_l, E_0\right) \equiv \left[\frac{1-\nu}{\nu}\frac{\mathcal{C}_l(N)}{E_0}\right]^{\nu} - 1$. **Q.E.D.**

1.A.5 Proof of Proposition 1.2

For the first part of the proposition, note that $i^d = i - s^d$ and s^d is decreasing in $\mathcal{C}_l$. It is straightforward that i^d is increasing in $\mathcal{C}_l$. For the second part of the proposition, take the

second-order cross derivative of s^d with respect to i and $\mathcal{C}_l$:

$$\begin{aligned}\frac{\partial^2 s^d}{\partial i \partial \mathcal{C}_l} &= \frac{\psi\theta_d}{\left[\mathcal{C}_l\left(1+i\right)^{-\frac{1-\nu}{\nu}}+\theta_d+\left(1+i\right)E_0\right]^2}\frac{1-\nu}{\nu}\left(1+i\right)^{-\frac{1}{\nu}} \\ &\quad -\frac{2\psi\theta_d}{\left[\mathcal{C}_l\left(1+i\right)^{-\frac{1-\nu}{\nu}}+\theta_d+\left(1+i\right)E_0\right]^3}\left\{\frac{1-\nu}{\nu}\mathcal{C}_l\left(1+i\right)^{-\frac{1}{\nu}}-E_0\right\}\left(1+i\right)^{-\frac{1-\nu}{\nu}}.\end{aligned}$$

For $\frac{\partial^2 i^d}{\partial i \partial \mathcal{C}_l} > 0$ it is equivalent to prove $\frac{\partial^2 s^d}{\partial i \partial \mathcal{C}_l} < 0$. This condition holds if and only if

$$\left(1-\nu\right)\theta_d+\left(1+i\right)E_0\left(1+\nu\right) < \left(1-\nu\right)\mathcal{C}_l\left(1+i\right)^{-\frac{1-\nu}{\nu}} \Leftrightarrow i < i_1\left(\mathcal{C}_l, E_0\right). \tag{1.48}$$

Note that the left-hand side of (1.48) is increasing in i and increases to infinity, and the right-hand side is decreasing in i and decreases to 0, then there exists a unique value $i_1\left(\mathcal{C}_l, E_0\right)$ below which the above inequality holds. The threshold value $i_1\left(\mathcal{C}_l, E_0\right)$ has following properties. First, it is smaller than $i_0\left(\mathcal{C}_l, E_0\right)$, since $\frac{\partial^2 s^d}{\partial i \partial \mathcal{C}_l}\Big|_{i=i_0(\mathcal{C}_l,E_0)} > 0$. Second, if $\left(1-\nu\right)\theta_d + E_0\left(1+\nu\right) < \left(1-\nu\right)\mathcal{C}_l$, then $i_1\left(\mathcal{C}_l, E_0\right)$ is greater than zero. Third, since the right-hand side of (1.48) is increasing in $\mathcal{C}_l$, and the left-hand side is increasing in E_0, by implicit function theorem we can get that $i_1\left(\mathcal{C}_l, E_0\right)$ increases in $\mathcal{C}_l$ and decreases in E_0. **Q.E.D.**

1.A.6 Proof of Proposition 1.3

The second-order derivative of s^d with respect to i is

$$\begin{aligned}\frac{\partial^2 s^d}{\partial i^2} &= \frac{\psi\theta_d}{\left[\mathcal{C}_l\left(1+i\right)^{-\frac{1-\nu}{\nu}}+\theta_d+\left(1+i-\psi\right)E_0\right]^2}\left\{-\frac{\left(1-\nu\right)\mathcal{C}_l}{\nu^2}\left(1+i\right)^{-\frac{1}{\nu}-1}\right\} \\ &\quad +\frac{2\psi\theta_d}{\left[\mathcal{C}_l\left(1+i\right)^{-\frac{1-\nu}{\nu}}+\theta_d+\left(1+i-\psi\right)E_0\right]^3}\left\{\frac{1-\nu}{\nu}\mathcal{C}_l\left(1+i\right)^{-\frac{1}{\nu}}-E_0\right\}^2.\end{aligned}$$

It implies that $\frac{\partial^2 s^d}{\partial i^2} > 0$ if and only if function $G\left(i; \mathcal{C}_l, E_0\right) > 0$, where

$$G\left(i; \mathcal{C}_l, E_0\right) \equiv \frac{\left(1-\nu\right)\left(1-2\nu\right)}{\nu^2} \frac{\mathcal{C}_l}{\left(1+i\right)^{\frac{1}{\nu}}} + 2E_0^2 \frac{\left(1+i\right)^{\frac{1}{\nu}}}{\mathcal{C}_l} - \frac{1-\nu}{\nu^2}\left[\frac{\theta_d - \psi E_0}{1+i} + \left(1+4\nu\right)E_0\right].$$

The function $G\left(i\right)$ has following properties. First, we have $G\left(i_0\left(\mathcal{C}_l, E_0\right); \mathcal{C}_l, E_0\right) < 0$. This is directly proved from that $\frac{\partial^2 s^d}{\partial i^2}\Big|_{i=i_0(\mathcal{C}_l, E_0)} < 0$. Second, $G\left(i; \mathcal{C}_l, E_0\right)$ is a U-shape function over $i \geq -1$. This is proved by taking first-order derivative of G with respect to i. One can show that the first-order derivative is positive if and only if

$$\frac{2E_0^2}{\nu \mathcal{C}_l}\left(1+i\right)^{\frac{1}{\nu}+1} + \frac{1-\nu}{\nu^2}\left(\theta_d - \psi E_0\right) > \frac{\left(1-\nu\right)\left(1-2\nu\right)}{\nu^3}\mathcal{C}_l\left(1+i\right)^{-\frac{1-\nu}{\nu}}.$$

In this inequality, the left-hand side is increasing in i and is positive at $i = -1$. Since $\nu < \frac{1}{2}$, the right-hand side decreases from infinite to zero as i increases from -1 to infinite. Thus there exists a unique value of i below (above) which G is decreasing (increasing) in i.

Third, $G\left(i; \mathcal{C}_l, E_0\right)$ is increasing $\mathcal{C}_l$ for any $i < \left(\frac{\mathcal{C}_l}{\nu E_0}\right)^{\nu}\left[\frac{\left(1-\nu\right)\left(1-2\nu\right)}{2}\right]^{\frac{\nu}{2}} - 1$. The threshold value is given by taking the first-order derivative of G with respect to $\mathcal{C}_l$ and letting the derivative to be positive. Moreover, the value of G at this threshold is negative, i.e.

$$\begin{aligned}
& G\left(i = \left(\frac{\mathcal{C}_l}{\nu E_0}\right)^{\nu}\left[\frac{\left(1-\nu\right)\left(1-2\nu\right)}{2}\right]^{\frac{\nu}{2}} - 1; \mathcal{C}_l, E_0\right) \\
= & \frac{2E_0}{\nu}\sqrt{2\left(1-\nu\right)\left(1-2\nu\right)} - \frac{\left(1-\nu\right)\left(1+4\nu\right)}{\nu^2}E_0 - \frac{1-\nu}{\nu^2}\frac{\theta_d - \psi E_0}{1+i} \\
< & \frac{\left(1-\nu\right)E_0}{\nu}\left[2\sqrt{2\frac{1-2\nu}{1-\nu}} - \frac{1+4\nu}{\nu}\right] < \frac{\left(1-\nu\right)E_0}{\nu}\left(2\sqrt{2}-4\right) < 0.
\end{aligned}$$

This implies that (i) when $\mathcal{C}_l \leq \nu E_0\left[\frac{2}{(1-\nu)(1-2\nu)}\right]^{\frac{1}{2}}$, $G\left(i; \mathcal{C}_l, E_0\right)$ is negative for any $i \in [0, i_0\left(\mathcal{C}_l, E_0\right)]$; (ii) when $\mathcal{C}_l > \nu E_0\left[\frac{2}{(1-\nu)(1-2\nu)}\right]^{\frac{1}{2}}$, $G\left(0; \mathcal{C}_l, E_0\right)$ is increasing in $\mathcal{C}_l$. It implies that there exists a unique threshold value $\widehat{\mathcal{C}}_l\left(E_0\right)$, which is larger than $\nu E_0\left[\frac{2}{(1-\nu)(1-2\nu)}\right]^{\frac{1}{2}}$, such that $G\left(0; \mathcal{C}_l, E_0\right) > 0$ if and only if $\mathcal{C}_l > \widehat{\mathcal{C}}_l\left(E_0\right)$. The threshold value $\widehat{\mathcal{C}}_l\left(E_0\right)$ is increasing

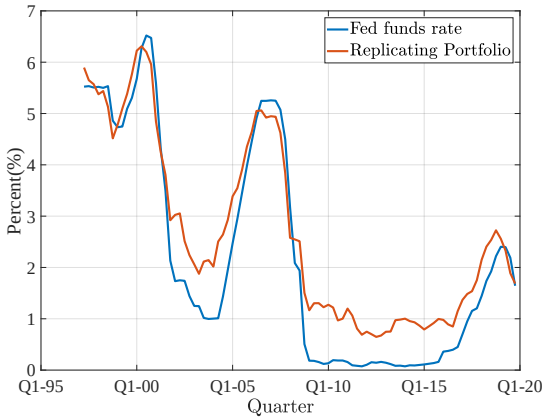
in E_0, because $G(0;\mathcal{C}_l,E_0)$ is decreasing in E_0 for any $\mathcal{C}_l > \nu E_0\left[\frac{2}{(1-\nu)(1-2\nu)}\right]^{\frac{1}{2}}$.

Finally, since G is a U-shape function in i, then there exists a unique threshold value $i_2(\mathcal{C}_l,E_0)$, which is smaller than $i_0(\mathcal{C}_l,E_0)$, such that $G(i;\mathcal{C}_l,E_0) > 0$ if $i < i_2(\mathcal{C}_l,E_0)$, and $G(i;\mathcal{C}_l,E_0) < 0$ if $i \in (i_2(\mathcal{C}_l,E_0), i_0(\mathcal{C}_l,E_0))$. The threshold value $i_2(\mathcal{C}_l,E_0)$ is positive if and only if $\mathcal{C}_l > \widehat{\mathcal{C}}_l(E_0)$. Moreover, whenever positive, $i_2(\mathcal{C}_l,E_0)$ is decreasing in E_0 and increasing in $\mathcal{C}_l$ and ψ, due to $\mathcal{C}_l > \nu E_0\left[\frac{2}{(1-\nu)(1-2\nu)}\right]^{\frac{1}{2}}$. **Q.E.D.**

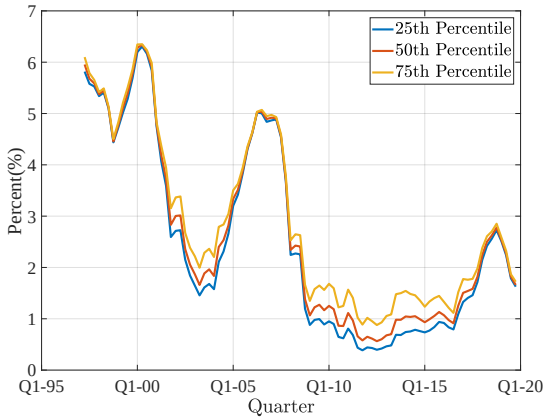
1.B Appendix: Figures

Figure 1.4: Replicating Portfolio

(a) Aggregate time series

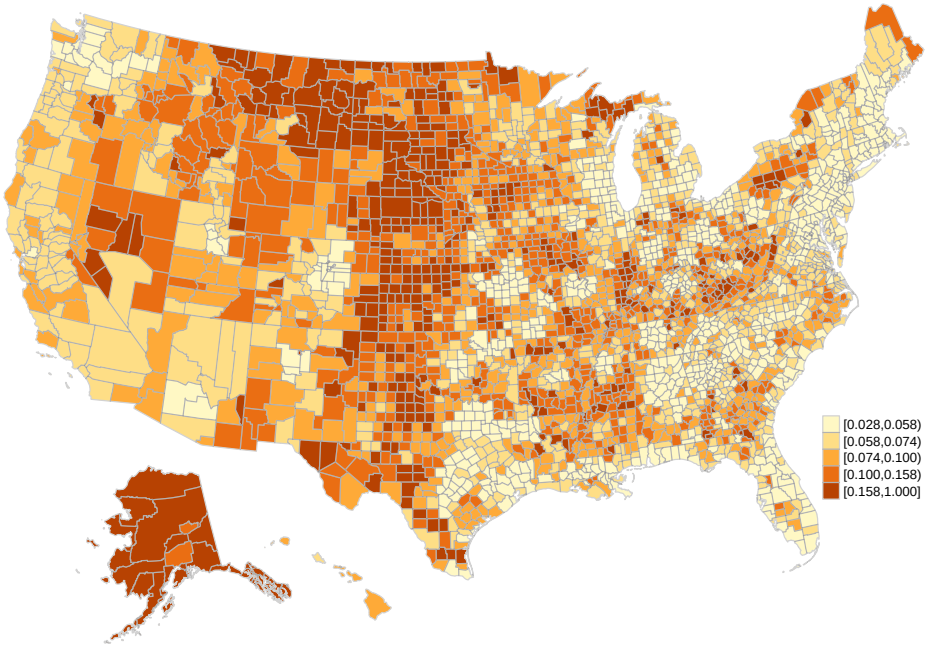


(b) Distribution of bank series



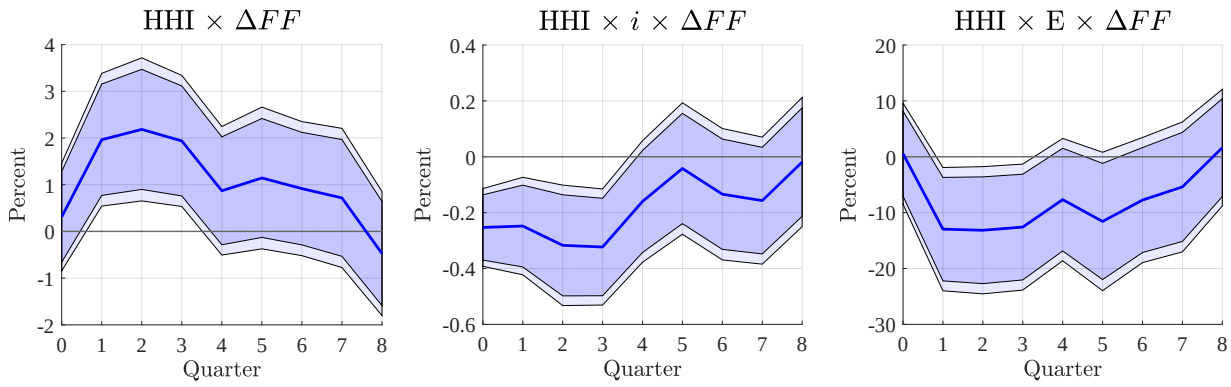
Note: This figure plots the time series of the yield rate on the replicating treasury portfolio. This portfolio replicates the repricing maturity structure of U.S. banks using the Call Reports data. Panel (a) plots the sequence of the aggregate replicating portfolio yield rate and the contemporaneous effective Fed funds rate. Panel (b) plots the quartiles of the yield rates on individual bank's replicating portfolio. The underlying data are from FRED and the Call Reports. The sample period is 1997Q1 to 2019Q4.

Figure 1.5: Home Mortgage Loan HHI in Local Banking Markets



Notes: This figure plots the yearly average Herfindahl index (HHI) of home mortgage loans by U.S. county. The HHI is calculated each year using the home mortgage loan market shares of all financial institutions that issue or purchase home mortgage loans in a given county, and then averaged over the period from 2000 to 2019. The underlying data are from the HMDA. The threshold values of colorbar in each panel are the quintiles of the HHI distribution.

Figure 1.6: Cumulative response of deposit spread at Bank



Notes: This figure plots the dynamic responses of bank deposit rates to interest rate shocks through the impact of loan market concentration and nominal interest rate. The blue line plots the estimated coefficients $\{\beta_{1,h}, \beta_{2,h}, \beta_{3,h}\}$ of equation (1.46) for horizon $h = 0, 1, 2, ..., 8$ (quarters). The estimation uses the local projection method of Jordà (2005). 95 and 90 percent confidence intervals are plotted using the standard errors clustered by banks. The left panel plots the sequence of $\beta_{1,h}$, the middle panel plots the sequence of $\beta_{2,h}$, and the right panel plots the sequence of $\beta_{3,h}$. The underlying data are from the Call Reports, HMDA, FRED, and Nakamura and Steinsson (2018). The sample period is from 1997Q1 to 2019Q4,

Figure 1.7: Cumulative response of bank balance sheet components

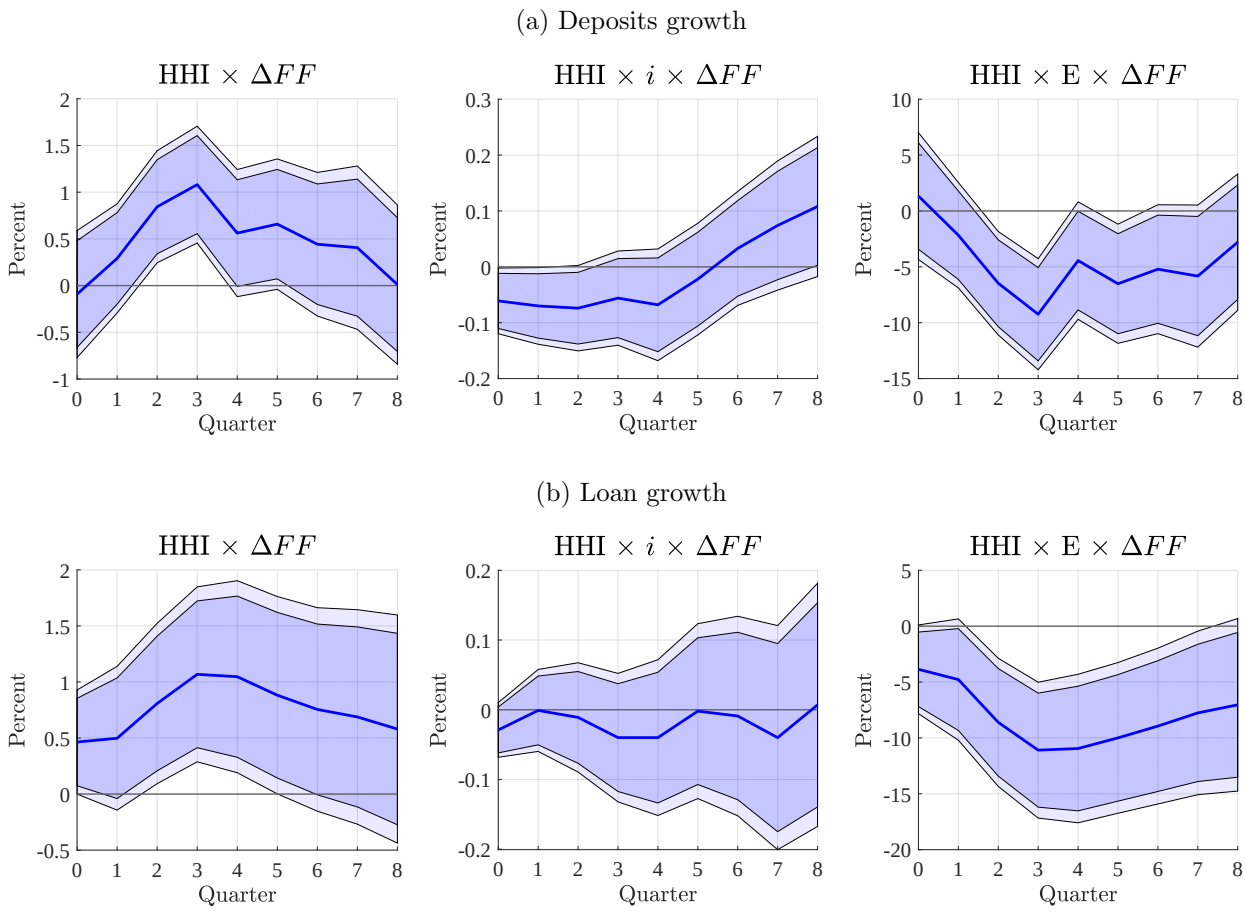
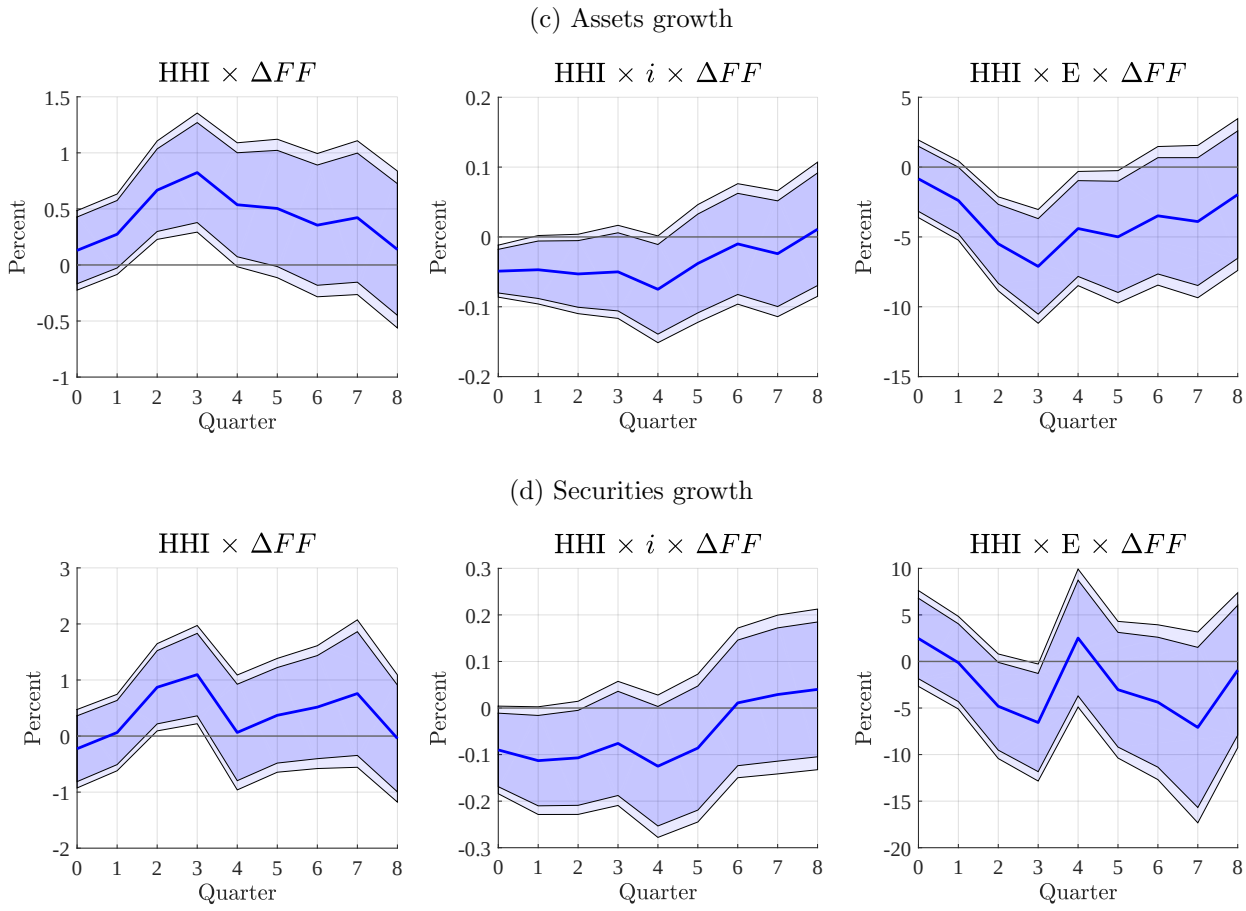
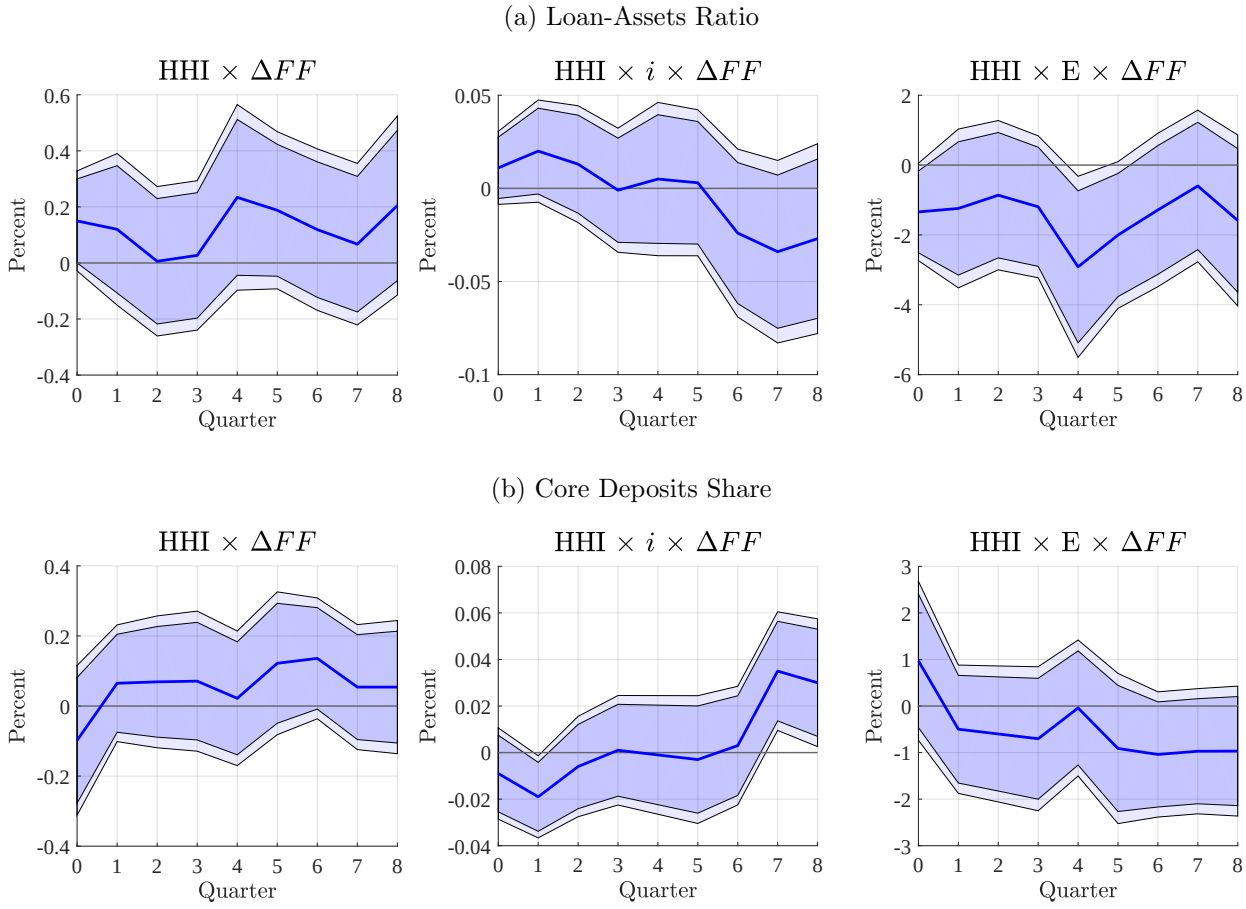


Figure 1.7: Cumulative response of bank balance sheet components (Cont.)



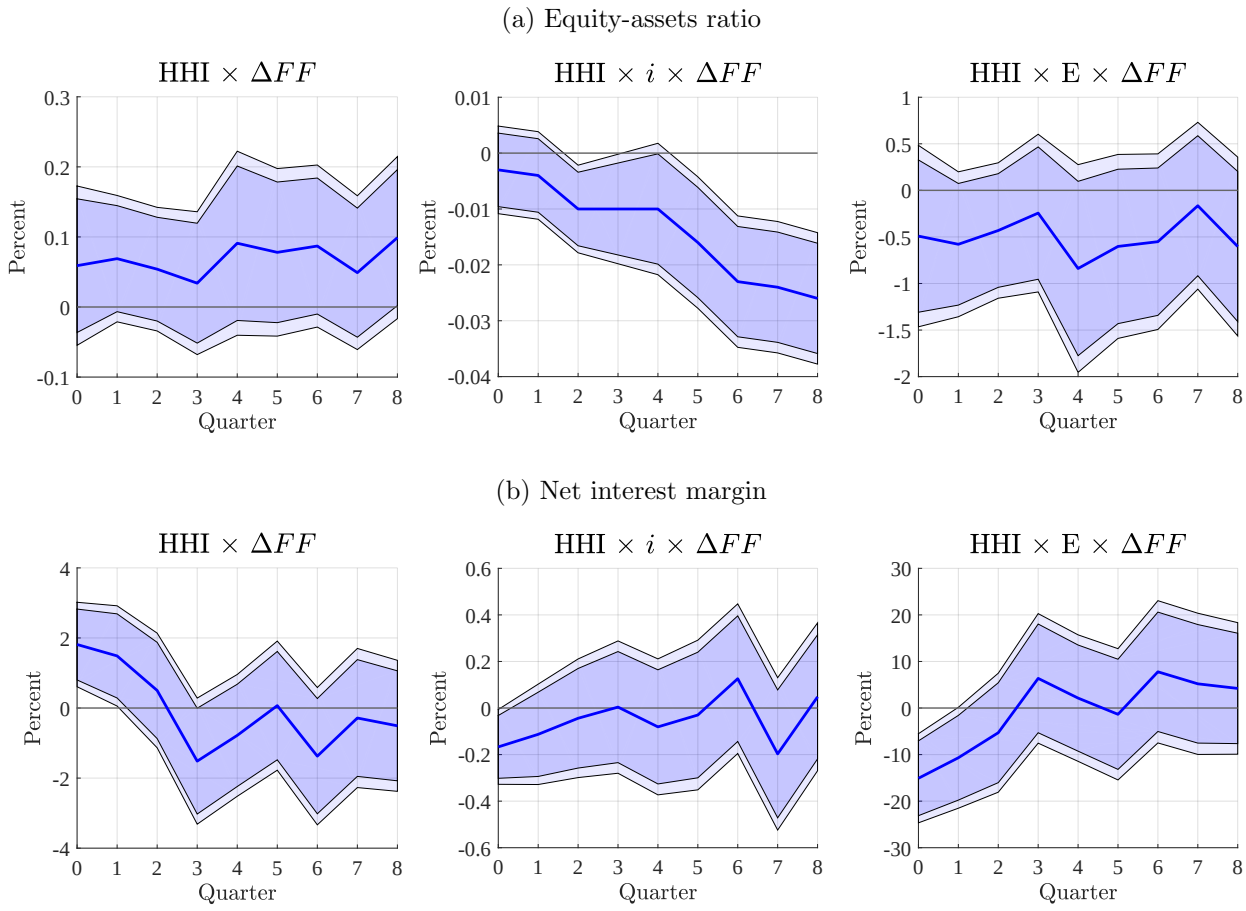
Notes: This figure plots the dynamic responses of the growth bank balance sheet components to interest rate shocks through the impact of loan market concentration and nominal interest rate. The blue line plots the estimated coefficients $\{\beta_{1,h}, \beta_{2,h}, \beta_{3,h}\}$ of equation (1.46) for horizon $h = 0, 1, 2, ..., 8$ (quarters). The estimation uses the local projection method of Jordà (2005). 95 and 90 percent confidence intervals are plotted using the standard errors clustered by banks. In each panel, the left figure plots the sequence of $\beta_{1,h}$, the middle figure plots the sequence of $\beta_{2,h}$, and the right figure plots the sequence of $\beta_{3,h}$. The underlying data are from the Call Reports, HMDA, FRED, and Nakamura and Steinsson (2018). The sample period is from 1997Q1 to 2019Q4,

Figure 1.8: Cumulative response of bank balance sheet structure



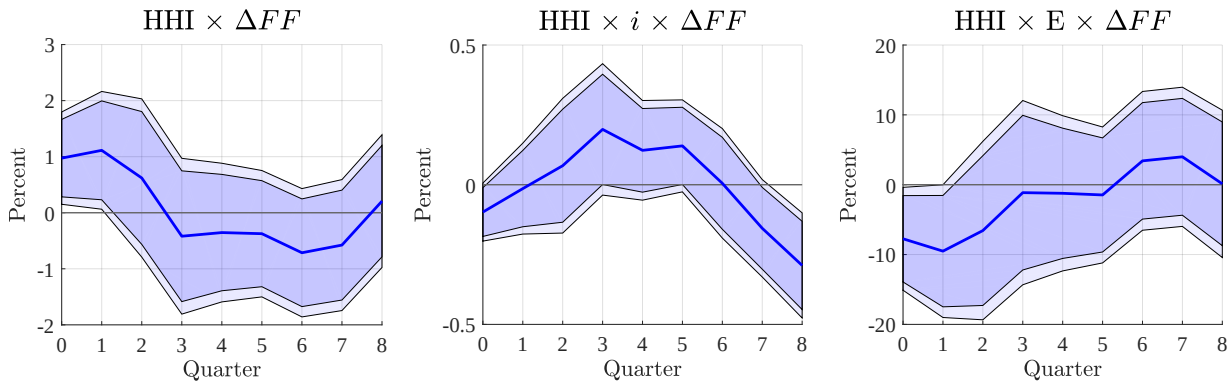
Notes: This figure plots the dynamic responses of bank balance sheet structure to interest rate shocks through the impact of loan market concentration and nominal interest rate. The blue line plots the estimated coefficients $\{\beta_{1,h}, \beta_{2,h}, \beta_{3,h}\}$ of equation (1.46) for horizon $h = 0, 1, 2, ..., 8$ (quarters). The estimation uses the local projection method of Jordà (2005). 95 and 90 percent confidence intervals are plotted using the standard errors clustered by banks. In each panel, the left figure plots the sequence of $\beta_{1,h}$, the middle figure plots the sequence of $\beta_{2,h}$, and the right figure plots the sequence of $\beta_{3,h}$. The underlying data are from the Call Reports, HMDA, FRED, and Nakamura and Steinsson (2018). The sample period is from 1997Q1 to 2019Q4,

Figure 1.9: Cumulative response of bank profitability



Notes: This figure plots the dynamic responses of bank profitability measures to interest rate shocks through the impact of loan market concentration and nominal interest rate. The blue line plots the estimated coefficients $\{\beta_{1,h}, \beta_{2,h}, \beta_{3,h}\}$ of equation (1.46) for horizon $h = 0, 1, 2, ..., 8$ (quarters). The estimation uses the local projection method of Jordà (2005). 95 and 90 percent confidence intervals are plotted using the standard errors clustered by banks. In each panel, the left figure plots the sequence of $\beta_{1,h}$, the middle figure plots the sequence of $\beta_{2,h}$, and the right figure plots the sequence of $\beta_{3,h}$. The underlying data are from the Call Reports, HMDA, FRED, and Nakamura and Steinsson (2018). The sample period is from 1997Q1 to 2019Q4,

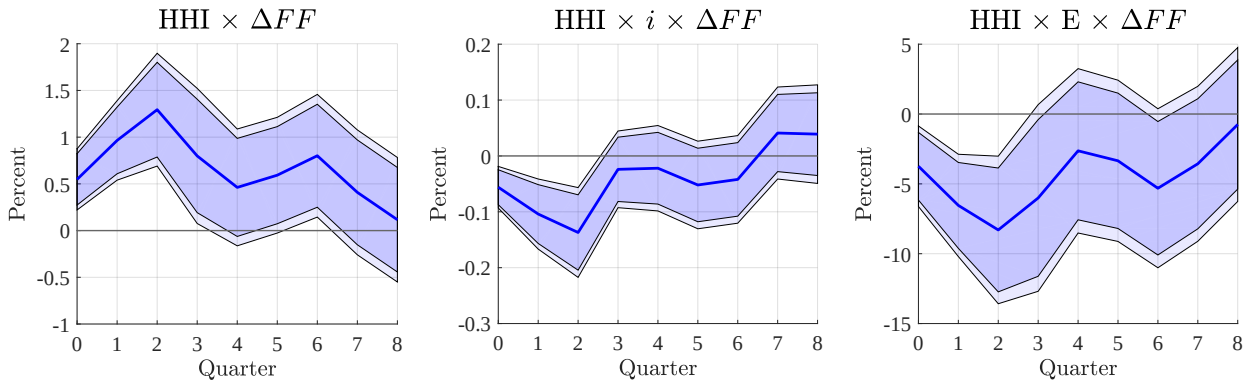
Figure 1.10: Cumulative response of deposit spread at Bank



Notes: This figure plots the dynamic responses of bank deposit rates to interest rate shocks through the impact of loan market concentration and nominal interest rates. The blue line plots the estimated coefficients $\{\beta_{1,h}, \beta_{2,h}, \beta_{3,h}\}$ of equation (1.46) for horizon $h = 0, 1, 2, ..., 8$ (quarters). The estimation uses the local projection method of Jordà (2005). 95 and 90 percent confidence intervals are plotted using the standard errors clustered by banks. The left panel plots the sequence of $\beta_{1,h}$, the middle panel plots the sequence of $\beta_{2,h}$, and the right panel plots the sequence of $\beta_{3,h}$. The underlying data are from the Call Reports, HMDA, FRED, and Jarociński and Karadi (2020). The sample period is from 1997Q1 to 2019Q4,

Figure 1.11: Cumulative response of bank balance sheet components

(a) Deposits growth



(b) Loan growth

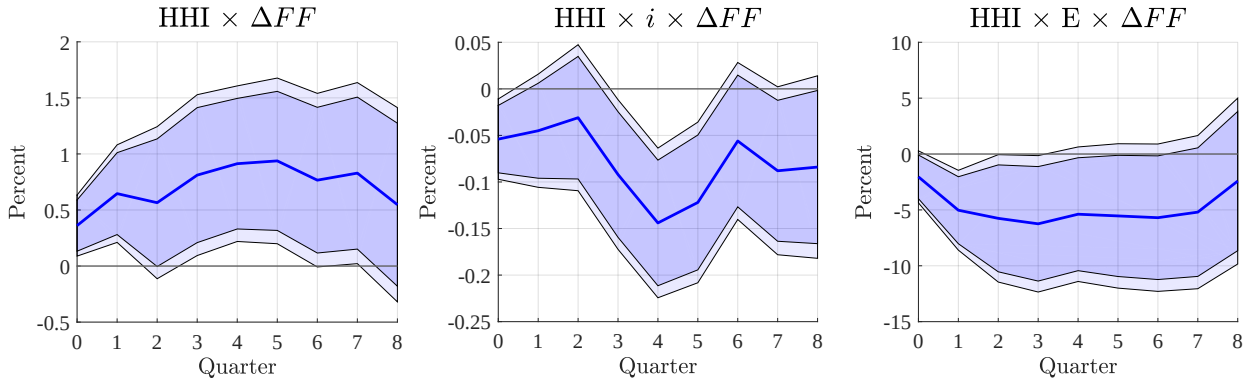
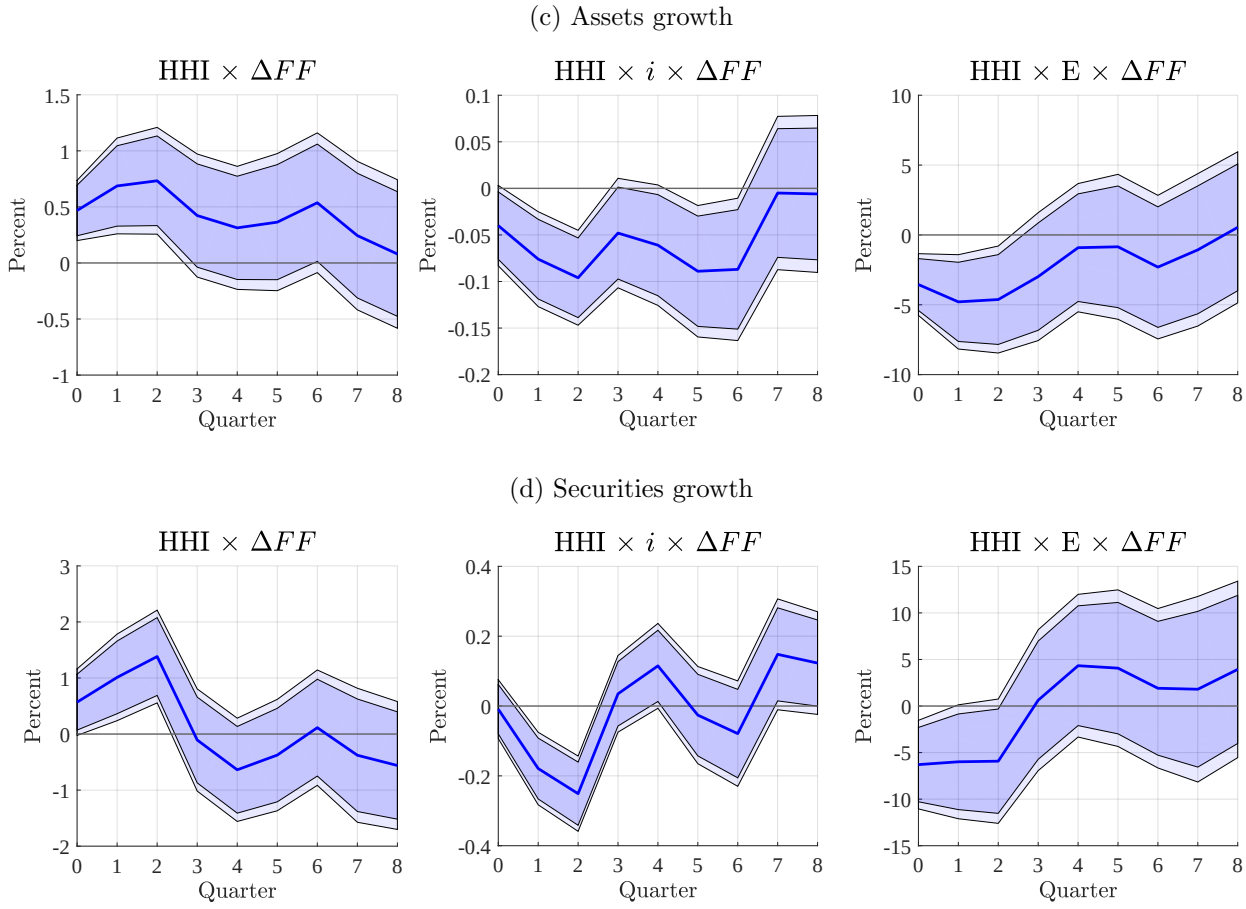
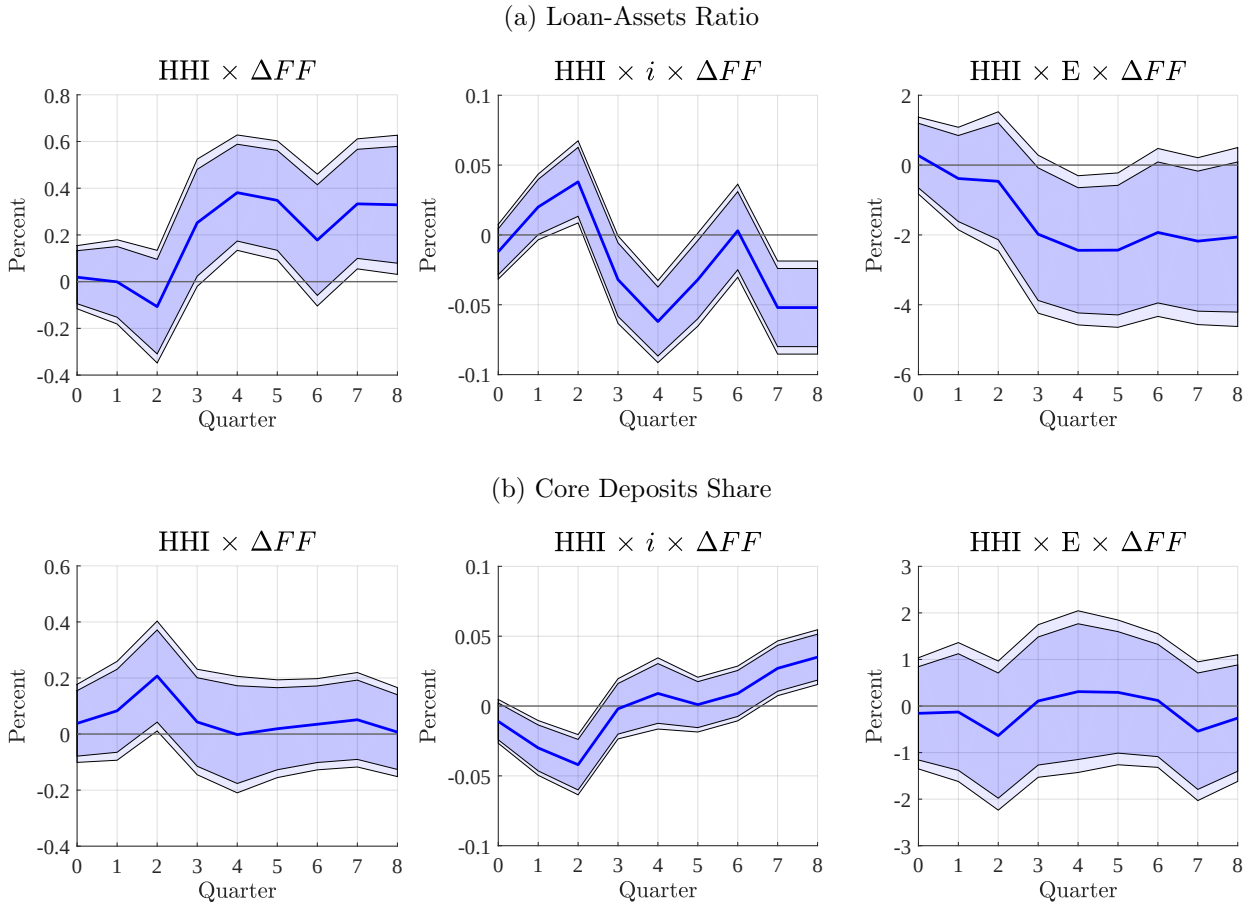


Figure 1.11: Cumulative response of bank balance sheet components (Cont.)



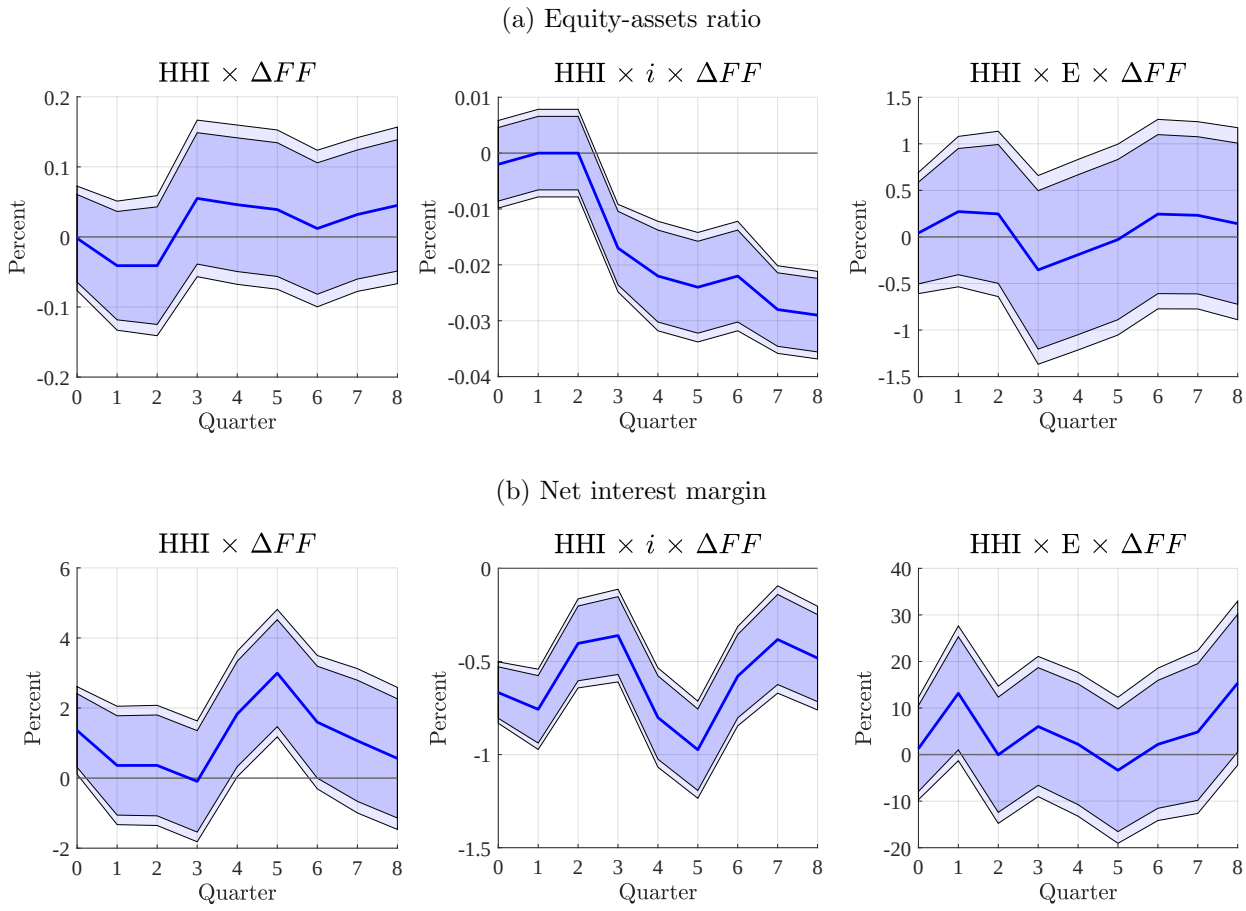
Notes: This figure plots the dynamic responses of the growth bank balance sheet components to interest rate shocks through the impact of loan market concentration and nominal interest rate. The blue line plots the estimated coefficients $\{\beta_{1,h}, \beta_{2,h}, \beta_{3,h}\}$ of equation (1.46) for horizon $h = 0, 1, 2, ..., 8$ (quarters). The estimation uses the local projection method of Jordà (2005). 95 and 90 percent confidence intervals are plotted using the standard errors clustered by banks. In each panel, the left figure plots the sequence of $\beta_{1,h}$, the middle figure plots the sequence of $\beta_{2,h}$, and the right figure plots the sequence of $\beta_{3,h}$. The underlying data are from the Call Reports, HMDA, FRED, and Jarociński and Karadi (2020). The sample period is from 1997Q1 to 2019Q4,

Figure 1.12: Cumulative response of bank balance sheet structure



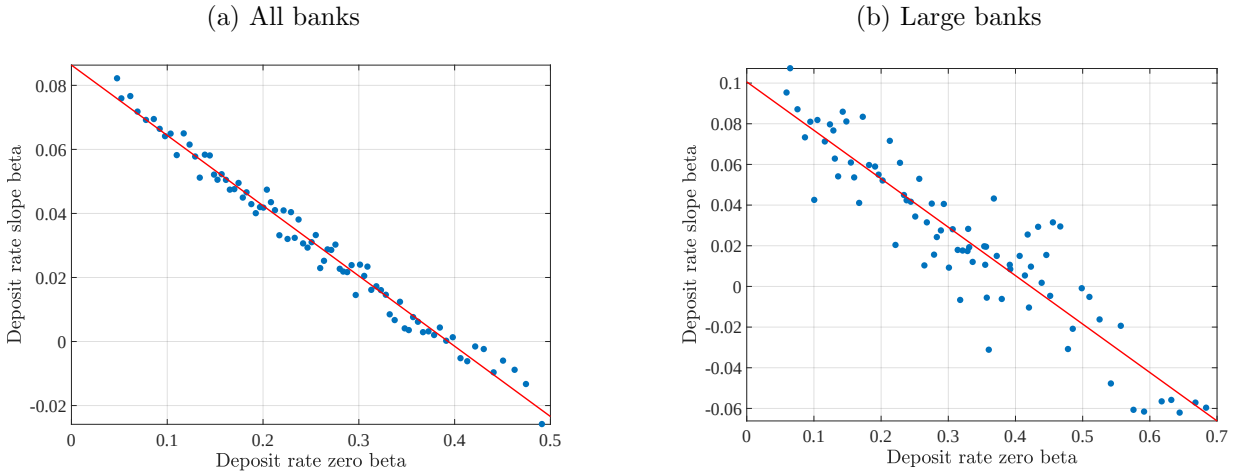
Notes: This figure plots the dynamic responses of bank balance sheet structure to interest rate shocks through the impact of loan market concentration and nominal interest rate. The blue line plots the estimated coefficients $\{\beta_{1,h}, \beta_{2,h}, \beta_{3,h}\}$ of equation (1.46) for horizon $h = 0, 1, 2, ..., 8$ (quarters). The estimation uses the local projection method of Jordà (2005). 95 and 90 percent confidence intervals are plotted using the standard errors clustered by banks. In each panel, the left figure plots the sequence of $\beta_{1,h}$, the middle figure plots the sequence of $\beta_{2,h}$, and the right figure plots the sequence of $\beta_{3,h}$. The underlying data are from the Call Reports, HMDA, FRED, and Jarociński and Karadi (2020). The sample period is from 1997Q1 to 2019Q4,

Figure 1.13: Cumulative response of bank profitability



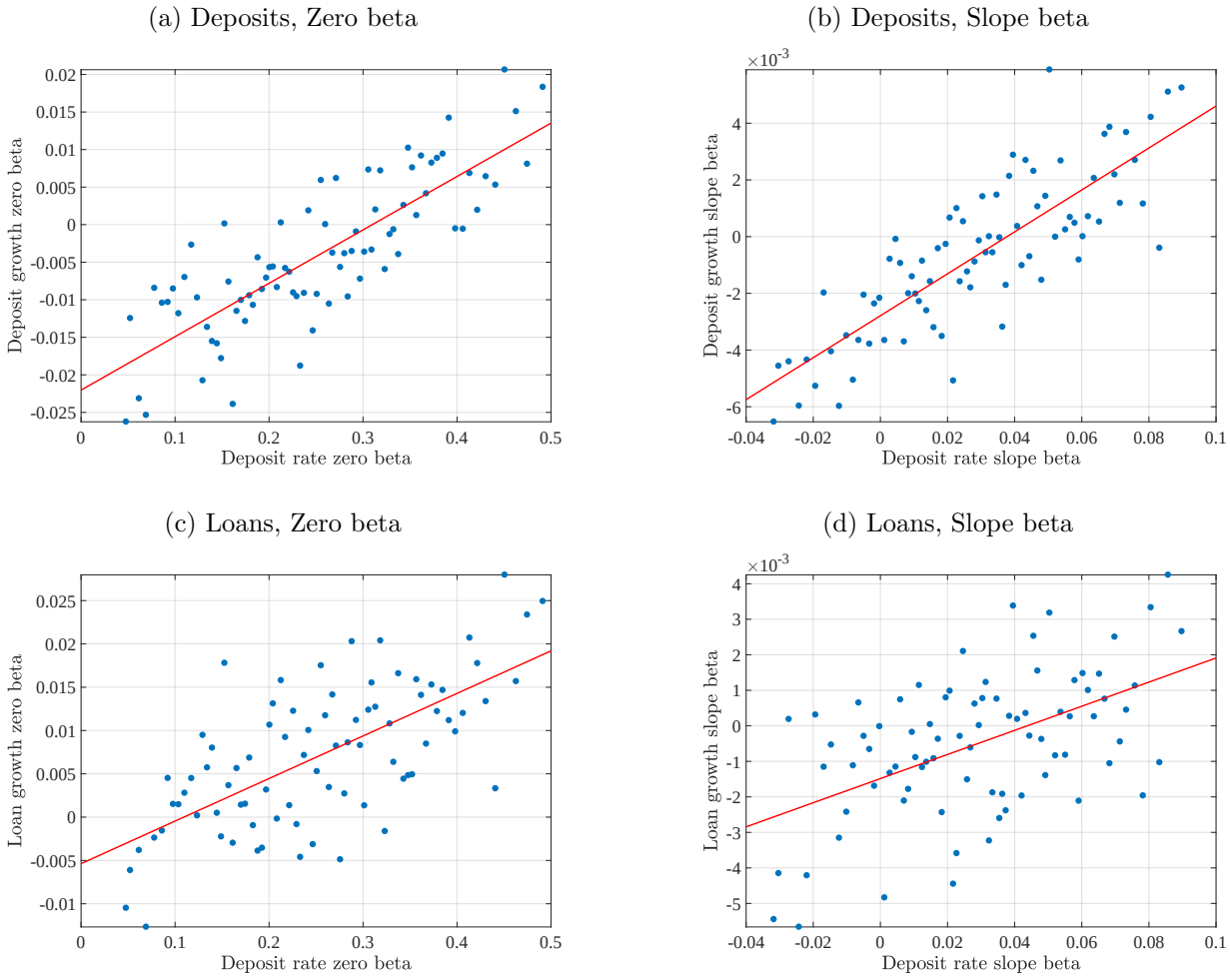
Notes: This figure plots the dynamic responses of bank profitability measures to interest rate shocks through the impact of loan market concentration and nominal interest rate. The blue line plots the estimated coefficients $\{\beta_{1,h}, \beta_{2,h}, \beta_{3,h}\}$ of equation (1.46) for horizon $h = 0, 1, 2, ..., 8$ (quarters). The estimation uses the local projection method of Jordà (2005). 95 and 90 percent confidence intervals are plotted using the standard errors clustered by banks. In each panel, the left figure plots the sequence of $\beta_{1,h}$, the middle figure plots the sequence of $\beta_{2,h}$, and the right figure plots the sequence of $\beta_{3,h}$. The underlying data are from the Call Reports, HMDA, FRED, and Jarociński and Karadi (2020). The sample period is from 1997Q1 to 2019Q4,

Figure 1.14: General deposit spread betas



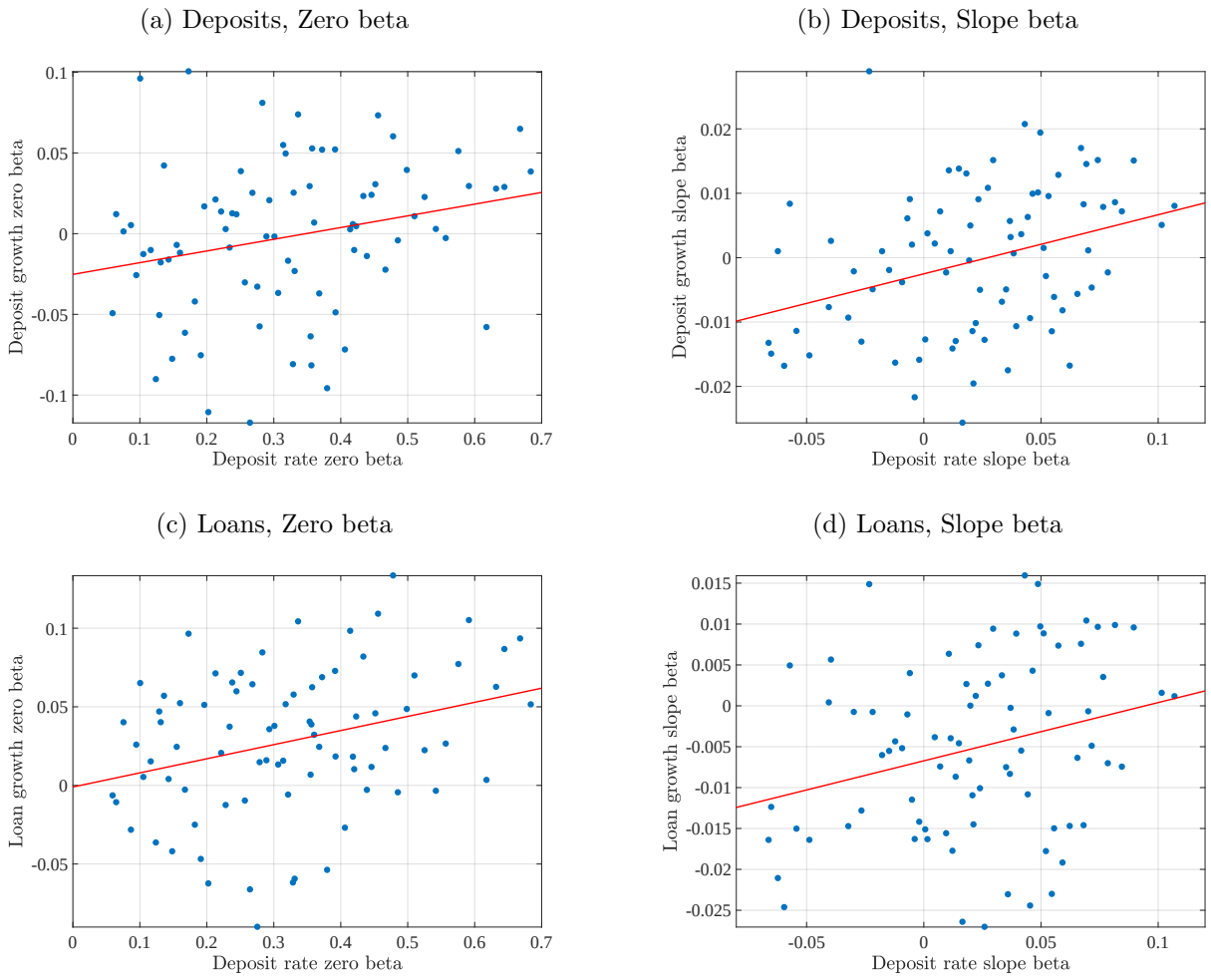
Notes: This figure shows scatter plots of average deposit rate slope betas over 100 bins of deposit rate zero beta. The zero beta measures the passthrough of Fed funds rate to individual bank's deposit rate when the Fed funds rate is zero. The slope beta measures the change in the passthrough when the Fed funds rate increases by 100 bps. Only banks with at least 60 quarterly observations are included. Both betas are winsorized at 10%. The left panel plots the results for all banks, and the right panel plots the results for largest 5% banks by assets. The sample is from 1997Q1 to 2019Q4.

Figure 1.15: Flow betas vs deposit rate betas (all banks)



Notes: This figure shows scatter plots of average flow betas over 100 bins of deposit rate betas. Panel (a) is the bin scatter plot of deposit growth zero beta versus deposit rate zero beta. Panel (b) is the bin scatter plot of deposit growth slope beta versus deposit rate slope beta. Panel (c) is the bin scatter plot of loan growth zero beta versus deposit rate zero beta. Panel (d) is the bin scatter plot of loan growth slope beta versus deposit rate slope beta. The deposit rate zero beta measures the passthrough of Fed funds rate to a bank's deposit rate when the Fed funds rate is zero. The deposit rate slope beta measures the change in the passthrough when the Fed funds rate increases by 100 bps. The deposit (loan) growth zero beta measures the sensitivity of a bank's log deposits (loans) to the Fed funds rate when the Fed funds rate is zero. The deposit (loan) growth slope beta measures the change in the sensitivity when the Fed funds rate increases by 100 bps. Only banks with at least 60 quarterly observations are included. Both betas are winsorized at 10%. The panels plot the results for all banks. The sample is from 1997Q1 to 2019Q4.

Figure 1.16: Flow betas vs deposit rate betas (large banks)



Notes: This figure shows scatter plots of average flow betas over 100 bins of deposit rate betas. Panel (a) is the bin scatter plot of deposit growth zero beta versus deposit rate zero beta. Panel (b) is the bin scatter plot of deposit growth slope beta versus deposit rate slope beta. Panel (c) is the bin scatter plot of loan growth zero beta versus deposit rate zero beta. Panel (d) is the bin scatter plot of loan growth slope beta versus deposit rate slope beta. The deposit rate zero beta measures the passthrough of Fed funds rate to a bank's deposit rate when the Fed funds rate is zero. The deposit rate slope beta measures the change in the passthrough when the Fed funds rate increases by 100 bps. The deposit (loan) growth zero beta measures the sensitivity of a bank's log deposits (loans) to the Fed funds rate when the Fed funds rate is zero. The deposit (loan) growth slope beta measures the change in the sensitivity when the Fed funds rate increases by 100 bps. Only banks with at least 60 quarterly observations are included. Both betas are winsorized at 10%. The panels plot the results for largest 5% banks by assets. The sample is from 1997Q1 to 2019Q4.

1.C Appendix: Tables

Table 1.2: Summary statistics

	All		< 2010		≥ 2010	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Panel A: Branch Deposit Rates (Ratewacth)						
ΔRate (MM, %)	-0.04	0.21	-0.07	0.30	-0.01	0.07
ΔRate (3M CD, %)	-0.05	0.27	-0.09	0.38	-0.02	0.08
ΔRate (6M CD, %)	-0.05	0.30	-0.09	0.42	-0.02	0.11
ΔRate (12M CD, %)	-0.05	0.31	-0.09	0.42	-0.02	0.14
Obs. (product×branch×quarter)	2,883,416		1,381,701		1,501,715	
Panel B: Branch deposits (FDIC)						
Deposit growth (%)	7.66	26.66	8.94	29.54	6.09	22.57
Obs. (branch×year)	1,944,437		1,068,767		875,670	
Panel C: County characterstics (FDIC and HMDA)						
Branch-HMDA-HHI	0.10	0.11	0.08	0.11	0.12	0.10
Obs. (counties)	3,225		3,219		3,216	
Panel D: Bank characteristics (Call Reports)						
Rate (deposits, %)	-0.04	0.32	-0.05	0.39	-0.02	0.14
Net interest margin (%)	3.69	0.85	3.84	0.85	3.44	0.77
Equity (%)	11.40	0.07	11.23	0.07	11.71	0.06
Assets (mill. $)	1,503	29,235	967	17,640	2,448	42,558
Bank-HMDA-HHI	0.06	0.03	0.04	0.03	0.07	0.03
# of banks	12,950		12,263		7,808	
Obs. (bank×quarter)	699,744		446,418		253,326	
Panel E: Aggregate series						
Replicating portfolio yield rate (%)	2.70	1.78	3.83	1.56	1.25	0.59
Obs. (quarter)	91		51		40	

Notes: This table provides summary statistics at the branch, bank, county and aggregate levels. All panels provide a breakdown by subperiods over the quarters up to 2009Q4 and after 2010Q1. Panel A presents data on the quarterly changes of branch-level deposit rates of four deposit products. "MM" represents the 25K Money Market account. "3M CD", "6M CD" and "12M CD" represent the 10K CD accounts with 3-month, 6-month and 12-month maturity. The underlying data are from Ratewatch from January 2001 to December 2019. Panel B presents data on the annual growth of branch deposit volumes. The underlying data are from FDIC from 1994 to 2019. Panel C presents data on a county's Herfindahl index of home mortgage loans. The underlying data are from HMDA from 2000 to 2019. Panel D presents data on bank characteristics. The underlying data are from the Call Reports from 1997Q1 to 2019 Q4. Panel E presents data on the yield rate of the aggregate replicating treasury portfolio. The underlying data are from the Call Reports and FRED from 1997Q1 to 2019Q4.

Table 1.3: Identification of the channel: preliminary results on deposit rates

Dependent Variable:	Δbranch deposit rate (quarterly)							
Loan market power:	Bank-HMDA-HHI							
Deposit product:	Money Market Account				3M CD			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$1\{t < 2010Q1\} \times HHI \times \Delta i$	0.174	0.131	0.303	0.299	-0.601**	-0.472*	-0.047	-0.042
	(0.258)	(0.252)	(0.223)	(0.223)	(0.251)	(0.246)	(0.216)	(0.216)
$1\{t \geq 2010Q1\} \times HHI \times \Delta i$	0.205	0.209	0.123	0.141	0.378**	0.543***	0.400***	0.406***
	(0.166)	(0.162)	(0.110)	(0.109)	(0.186)	(0.183)	(0.153)	(0.152)
Obs	206,211	206,211	206,211	206,211	195,515	195,515	195,515	195,515
Adj R^2	0.789	0.788	0.782	0.787	0.708	0.707	0.685	0.694
Deposit product:	6M CD				12M CD			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
$1\{t < 2010Q1\} \times HHI \times \Delta i$	-0.895***	-0.797***	-0.526***	-0.527***	-0.321*	-0.269	-0.484***	-0.494***
	(0.234)	(0.228)	(0.196)	(0.196)	(0.183)	(0.183)	(0.176)	(0.175)
$1\{t \geq 2010Q1\} \times HHI \times \Delta i$	0.484**	0.631***	0.478***	0.485***	0.790***	0.956***	0.608***	0.620***
	(0.197)	(0.194)	(0.174)	(0.174)	(0.227)	(0.233)	(0.192)	(0.191)
Obs	213,701	213,701	213,701	213,701	214,374	214,374	214,374	214,374
Adj R^2	0.669	0.668	0.641	0.650	0.670	0.668	0.639	0.648
Controls (all panels):								
Branch FE	Y	Y	Y	N	Y	Y	Y	N
Bank FE	Y	Y	Y	N	Y	Y	Y	N
County FE	N	N	Y	Y	N	N	Y	Y
Time FE	N	N	Y	Y	N	N	Y	Y
County×time FE	Y	Y	N	N	Y	Y	N	N
Bank controls	Y	N	Y	Y	Y	N	Y	Y

Notes: This table presents the estimation of equation (1.43) for branch deposit rates. The dependent variable is the quarterly change of a branch's deposit rate on a deposit product. For the independent variables listed in the table, $1\{t < 2010Q1\}$ and $1\{t \geq 2010Q1\}$ are the indicators of whether the observation is before 2010Q1 or after 2010Q1. HHI is the bank-level average Herfindahl index of home mortgage loans across counties (Bank-HMDA-HHI). Δi is the quarterly change in the Fed funds rate. The sample consists of all U.S. counties with branches of at least two different banks for identification. The deposit products include 25K money market accounts (Money Market Account) and 10K CD accounts with 3-month, 6-month and 12-month maturity (3M CD, 6M CD, 12M CD). The underlying data are from Ratewatch, Call Reports, HMDA and FRED. The sample period is from 2001Q1 to 2019Q4. Fixed effects are denoted at the bottom of the table. Standard errors clustered by county are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.4: Identification of the channel: preliminary results on deposit growth

Dependent Variable	$\Delta \log$ (Branch Deposits, Annual)			
Loan market power	Bank-HMDA-HHI			
	(1)	(2)	(3)	(4)
$1\{t < 2010Q1\} \times HHI \times \Delta i$	-0.070***	-0.075***	-0.005	0.002
	(0.026)	(0.026)	(0.017)	(0.017)
$1\{t \geq 2010Q1\} \times HHI \times \Delta i$	2.674***	2.093***	-0.407	-0.354
	(0.761)	(0.779)	(0.652)	(0.637)
Obs	1,284,427	1,284,427	1,284,427	1,284,427
Adj R^2	0.264	0.262	0.230	0.067
Controls (all panels):				
Branch FE	Y	Y	Y	N
Bank FE	Y	Y	Y	N
County	N	N	Y	Y
Time FE	N	N	Y	Y
County×time FE	Y	Y	N	N
Bank controls	Y	N	Y	Y

Notes: This table presents the estimation of equation (1.43) for branch deposit growth. The dependent variable is the annual log difference of a branch's total deposit volumes. For the independent variables listed in the table, $1\{t < 2010Q1\}$ and $1\{t \geq 2010Q1\}$ are the indicators of whether the observation is before 2010Q1 or after 2010Q1. HHI is the bank-level average Herfindahl index of home mortgage loans across counties (Bank-HMDA-HHI). Δi is the quarterly change in the Fed funds rate. The sample consists of all U.S. counties with branches of at least two different banks for identification. The underlying data are from FDIC, Call Reports, HMDA and FRED. The sample period is from 1997 to 2019. Fixed effects are denoted at the bottom of the table. Standard errors clustered by county are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.5: Identification of the channel: baseline results for deposit rates

Dependent Variable:	Δbranch deposit rate (quarterly)					
Loan market power:	Bank-HMDA-HHI					
Deposit product:	Money Market Account			3M CD		
	(1)	(2)	(3)	(4)	(5)	(6)
β_1: $HHI \times \Delta i$	0.120	-0.198	2.220***	-0.174	1.375***	2.736***
	(0.226)	(0.477)	(0.753)	(0.218)	(0.464)	(0.887)
β_2: $HHI \times i \times \Delta i$		0.103	-0.130		-0.489***	-0.520***
		(0.146)	(0.206)		(0.171)	(0.170)
β_3: $HHI \times E \times \Delta i$			-19.041***			-12.441*
			(6.525)			(6.813)
Obs	204,674	204,674	204,674	194,101	194,101	194,101
Adj R^2	0.988	0.988	0.988	0.973	0.973	0.973
Deposit product:	6M CD			12M CD		
	(7)	(8)	(9)	(10)	(11)	(12)
β_1: $HHI \times \Delta i$	-0.325	1.826***	2.308***	-0.269	2.010***	2.824***
	(0.201)	(0.441)	(0.695)	(0.190)	(0.452)	(0.635)
β_2: $HHI \times i \times \Delta i$		-0.658***	-0.677***		-0.682***	-0.705***
		(0.149)	(0.147)		(0.146)	(0.144)
β_3: $HHI \times E \times \Delta i$			-4.188			-7.116*
			(5.158)			(4.125)
Obs	212,418	212,418	212,418	212,824	212,824	212,824
Adj R^2	0.966	0.966	0.966	0.965	0.965	0.965
Controls (all panels):						
Branch FE	Y	Y	Y	Y	Y	Y
Bank FE	Y	Y	Y	Y	Y	Y
County×time FE	Y	Y	Y	Y	Y	Y
Bank controls	Y	Y	Y	Y	Y	Y

Notes: This table presents the estimation of equation (1.44) for branch deposit rates. The dependent variable is the quarterly change of a branch's deposit rate on a deposit product. For the independent variables listed in the table, HHI is the bank-level average Herfindahl index of home mortgage loans across counties (Bank-HMDA-HHI). i is the nominal interest rate. E is the bank equity-assets ratio. Δi is the quarterly change in the Fed funds rate. The sample consists of all U.S. counties with branches of at least two different banks for identification. The deposit products include 25K money market accounts (Money Market Account) and 10K CD accounts with 3-month, 6-month and 12-month maturity (3M CD, 6M CD, 12M CD). The underlying data are from Ratewatch, Call Reports, HMDA and FRED. The sample period is from 2001Q1 to 2019Q4. Fixed effects are denoted at the bottom of the table. Standard errors clustered by county are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.6: Robustness checks: alternative fixed effects and controls for the regressions of deposit rates

Dependent Variable:	Δbranch deposit rate (quarterly)							
Loan market power:	Bank-HMDA-HHI							
Deposit product:	Money Market Account				3M CD			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1: $HHI \times \Delta i$	-0.197	-0.272	0.007	0.025	1.400***	1.127***	0.990***	0.998***
	(0.477)	(0.238)	(0.224)	(0.223)	(0.464)	(0.313)	(0.289)	(0.287)
β_2: $HHI \times i \times \Delta i$	0.101	0.060	0.038	0.032	-0.497***	-0.383***	-0.269**	-0.271**
	(0.146)	(0.087)	(0.085)	(0.085)	(0.172)	(0.114)	(0.108)	(0.108)
Obs	204,674	204,674	204,674	204,674	194,101	194,101	194,101	194,101
Adj R^2	0.988	0.988	0.988	0.988	0.973	0.973	0.972	0.973
Deposit product:	6M CD				12M CD			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
β_1: $HHI \times \Delta i$	1.834***	1.198***	1.267***	1.275***	2.055***	1.488***	1.443***	1.482***
	(0.441)	(0.280)	(0.260)	(0.260)	(0.455)	(0.287)	(0.263)	(0.263)
β_2: $HHI \times i \times \Delta i$	-0.662***	-0.434***	-0.392***	-0.397***	-0.693***	-0.593***	-0.518***	-0.521***
	(0.150)	(0.100)	(0.095)	(0.096)	(0.148)	(0.104)	(0.097)	(0.097)
Obs	212,418	212,418	212,418	212,418	212,824	212,824	212,824	212,824
Adj R^2	0.966	0.965	0.964	0.964	0.965	0.964	0.963	0.963
Controls (all panels):								
Branch FE	Y	Y	Y	N	Y	Y	Y	N
Bank FE	Y	Y	Y	N	Y	Y	Y	N
County FE	N	Y	Y	Y	N	Y	Y	Y
Time FE	N	N	Y	Y	N	N	Y	Y
County×time FE	Y	N	N	N	Y	N	N	N
State×time FE	N	Y	N	N	N	Y	N	N
Bank controls	N	Y	Y	Y	N	Y	Y	Y

Notes: This table presents the estimation of equation (1.44) for branch deposit rates with alternative fixed effects and bank controls. In the estimation the interaction terms with bank equity-assets ratio are not included. The dependent variable is the quarterly change of a branch's deposit rate on a deposit product. For the independent variables listed in the table, HHI is the bank-level average Herfindahl index of home mortgage loans across counties (Bank-HMDA-HHI). i is the nominal interest rate. Δi is the quarterly change in the Fed funds rate. The sample consists of all U.S. counties with branches of at least two different banks for identification. The deposit products include 25K money market accounts and 10K CD accounts with 3-month, 6-month and 12-month maturity. The underlying data are from Ratewatch, Call Reports, HMDA and FRED. The sample period is from 2001Q1 to 2019Q4. Fixed effects are denoted at the bottom of the table. Standard errors clustered by county are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.7: Robustness checks: alternative fixed effects and controls for the regressions of deposit rates

Dependent Variable:	Δbranch deposit rate (quarterly)							
Loan market power:	Bank-HMDA-HHI							
Deposit product:	Money Market Account				3M CD			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1: $HHI \times \Delta i$	2.129***	1.622***	1.955***	1.995***	2.827***	1.911***	1.557***	1.487***
	(0.750)	(0.490)	(0.486)	(0.487)	(0.896)	(0.544)	(0.492)	(0.489)
β_2: $HHI \times i \times \Delta i$	0.033	0.005	-0.015	-0.025	-0.523***	-0.413***	-0.300***	-0.301***
	(0.146)	(0.087)	(0.085)	(0.085)	(0.171)	(0.114)	(0.107)	(0.107)
β_3: $HHI \times E \times \Delta i$	-20.678***	-17.006***	-17.583***	-17.676***	-13.252*	-6.895*	-4.720	-3.985
	(6.558)	(4.105)	(4.113)	(4.125)	(6.909)	(3.953)	(3.668)	(3.646)
Obs	204,674	204,674	204,674	204,674	194,101	194,101	194,101	194,101
Adj R^2	0.988	0.988	0.988	0.988	0.973	0.973	0.973	0.973
Deposit product:	6M CD				12M CD			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
β_1: $HHI \times \Delta i$	2.387***	1.773***	1.352***	1.226***	2.987***	2.420***	1.932***	1.836***
	(0.704)	(0.474)	(0.464)	(0.465)	(0.641)	(0.486)	(0.480)	(0.481)
β_2: $HHI \times i \times \Delta i$	-0.679***	-0.450***	-0.395***	-0.403***	-0.714***	-0.612***	-0.526***	-0.532***
	(0.147)	(0.099)	(0.094)	(0.095)	(0.146)	(0.104)	(0.096)	(0.097)
β_3: $HHI \times E \times \Delta i$	-4.897	-5.249	-0.919	0.481	-8.344**	-8.511**	-4.553	-3.161
	(5.243)	(3.393)	(3.443)	(3.468)	(4.195)	(3.401)	(3.507)	(3.529)
Obs	212,418	212,418	212,418	212,418	212,824	212,824	212,824	212,824
Adj R^2	0.966	0.965	0.964	0.964	0.965	0.964	0.963	0.963
Controls (all panels):								
Branch FE	Y	Y	Y	N	Y	Y	Y	N
Bank FE	Y	Y	Y	N	Y	Y	Y	N
County FE	N	Y	Y	Y	N	Y	Y	Y
Time FE	N	N	Y	Y	N	N	Y	Y
County×time FE	Y	N	N	N	Y	N	N	N
State×time FE	N	Y	N	N	N	Y	N	N
Bank controls	N	Y	Y	Y	N	Y	Y	Y

Notes: This table presents the estimation of equation (1.44) for branch deposit rates with alternative fixed effects and bank controls. In the estimation the interaction terms with bank equity-assets ratio are included. The dependent variable is the quarterly change of a branch's deposit rate on a deposit product. For the independent variables listed in the table, HHI is the bank-level average Herfindahl index of home mortgage loans across counties (Bank-HMDA-HHI). i is the nominal interest rate. E is the bank equity-assets ratio. Δi is the quarterly change in the Fed funds rate. The sample consists of all U.S. counties with branches of at least two different banks for identification. The deposit products include 25K money market accounts and 10K CD accounts with 3-month, 6-month and 12-month maturity. The underlying data are from Ratewatch, Call Reports, HMDA and FRED. The sample period is from 2001Q1 to 2019Q4. Fixed effects are denoted at the bottom of the table. Standard errors clustered by county are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.8: Robustness checks for deposit rate regressions: using Fed funds rate as the nominal rate

Dependent Variable:	Δbranch deposit rate (quarterly)							
Loan market power:	Bank-HMDA-HHI							
Deposit product:	MM		3M CD		6M CD		12M CD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1: $HHI \times \Delta i$	-0.030	2.091***	0.893***	2.298***	1.085***	1.584**	1.247***	2.025***
	(0.388)	(0.666)	(0.320)	(0.807)	(0.328)	(0.633)	(0.342)	(0.569)
β_2: $HHI \times i \times \Delta i$	0.052	0.015	-0.424***	-0.442***	-0.513***	-0.526***	-0.529***	-0.540***
	(0.122)	(0.121)	(0.137)	(0.138)	(0.120)	(0.120)	(0.120)	(0.120)
β_3: $HHI \times E \times \Delta i$		-19.892***		-13.081*		-4.363		-7.181*
		(6.643)		(6.881)		(5.141)		(4.157)
Obs	204,674	204,674	194,101	194,101	212,418	212,418	212,824	212,824
Adj R^2	0.988	0.988	0.973	0.973	0.966	0.966	0.965	0.965
Branch FE	Y	Y	Y	Y	Y	Y	Y	Y
Bank FE	Y	Y	Y	Y	Y	Y	Y	Y
County×time FE	Y	Y	Y	Y	Y	Y	Y	Y
Bank controls	Y	Y	Y	Y	Y	Y	Y	Y

Notes: This table presents the estimation of equation (1.44) for branch deposit rates, where the nominal interest rate is the Fed funds rate. In the estimation the interaction terms with bank equity-assets ratio are included. The dependent variable is the quarterly change of a branch's deposit rate on a deposit product. For the independent variables listed in the table, HHI is the bank-level average Herfindahl index of home mortgage loans across counties (Bank-HMDA-HHI). i is the nominal interest rate. E is the bank equity-assets ratio. Δi is the quarterly change in the Fed funds rate. The sample consists of all U.S. counties with branches of at least two different banks for identification. The deposit products include 25K money market accounts (MM) and 10K CD accounts with 3-month, 6-month and 12-month maturity (3M CD, 6M CD, 12M CD). The underlying data are from Ratewatch, Call Reports, HMDA and FRED. The sample period is from 2001Q1 to 2019Q4. Fixed effects are denoted at the bottom of the table. Standard errors clustered by county are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.9: Robustness checks for deposit rate regressions: using 1-year Treasury yield rate as the nominal rate

Dependent Variable:	Δbranch deposit rate (quarterly)							
Loan market power:	Bank-HMDA-HHI							
Deposit product:	MM		3M CD		6M CD		12M CD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1: $HHI \times \Delta i$	-0.089	2.049***	1.056***	2.473***	1.329***	1.847**	1.429***	2.239***
	(0.400)	(0.702)	(0.367)	(0.837)	(0.356)	(0.651)	(0.372)	(0.584)
β_2: $HHI \times i \times \Delta i$	0.078	0.034	-0.438***	-0.461***	-0.564***	-0.579***	-0.557***	-0.571***
	(0.130)	(0.131)	(0.152)	(0.152)	(0.131)	(0.130)	(0.130)	(0.129)
β_3: $HHI \times E \times \Delta i$		-19.822***		-13.084*		-4.543		-7.436*
		(6.623)		(6.904)		(5.174)		(4.143)
Obs	204,674	204,674	194,101	194,101	212,418	212,418	212,824	212,824
Adj R^2	0.988	0.988	0.973	0.973	0.966	0.966	0.965	0.965
Branch FE	Y	Y	Y	Y	Y	Y	Y	Y
Bank FE	Y	Y	Y	Y	Y	Y	Y	Y
County×time FE	Y	Y	Y	Y	Y	Y	Y	Y
Bank controls	Y	Y	Y	Y	Y	Y	Y	Y

Notes: This table presents the estimation of equation (1.44) for branch deposit rates, where the nominal interest rate is the 1-year Treasury yield rate. In the estimation the interaction terms with bank equity-assets ratio are included. The dependent variable is the quarterly change of a branch's deposit rate on a deposit product. For the independent variables listed in the table, HHI is the bank-level average Herfindahl index of home mortgage loans across counties (Bank-HMDA-HHI). i is the nominal interest rate. E is the bank equity-assets ratio. Δi is the quarterly change in the Fed funds rate. The sample consists of all U.S. counties with branches of at least two different banks for identification. The deposit products include 25K money market accounts (MM) and 10K CD accounts with 3-month, 6-month and 12-month maturity (3M CD, 6M CD, 12M CD). The underlying data are from Ratewatch, Call Reports, HMDA and FRED. The sample period is from 2001Q1 to 2019Q4. Fixed effects are denoted at the bottom of the table. Standard errors clustered by county are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.10: Robustness checks for deposit rate regressions: Pre-financial crisis results

Dependent Variable:	Δbranch deposit rate (quarterly)							
Loan market power:	Bank-HMDA-HHI							
Deposit product:	MM		3M CD		6M CD		12M CD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1: $HHI \times \Delta i$	-0.200	1.282***	2.405**	2.434	2.929***	3.577***	1.324	3.655***
	(1.036)	(1.372)	(0.951)	(1.708)	(0.836)	(1.282)	(0.868)	(1.248)
β_2: $HHI \times i \times \Delta i$	0.040	0.008	-0.865***	-0.816***	-0.950***	-0.965***	-0.687***	-0.675***
	(0.276)	(0.275)	(0.290)	(0.275)	(0.239)	(0.228)	(0.250)	(0.239)
β_3: $HHI \times E \times \Delta i$		-13.535		-2.557		-6.017		-23.586**
		(9.807)		(14.258)		(10.003)		(9.477)
Obs	65,104	65,104	61,773	61,773	67,250	67,250	67,300	67,300
Adj R^2	0.978	0.978	0.948	0.948	0.927	0.927	0.926	0.926
Branch FE	Y	Y	Y	Y	Y	Y	Y	Y
Bank FE	Y	Y	Y	Y	Y	Y	Y	Y
County×time FE	Y	Y	Y	Y	Y	Y	Y	Y
Bank controls	Y	Y	Y	Y	Y	Y	Y	Y

Notes: This table presents the estimation of equation (1.44) for branch deposit rates using the data before the 2008 global financial crisis (until June 2008). In the estimation the interaction terms with bank equity-assets ratio are included. The dependent variable is the quarterly change of a branch's deposit rate on a deposit product. For the independent variables listed in the table, HHI is the bank-level average Herfindahl index of home mortgage loans across counties (Bank-HMDA-HHI). i is the nominal interest rate. E is the bank equity-assets ratio. Δi is the quarterly change in the Fed funds rate. The sample consists of all U.S. counties with branches of at least two different banks for identification. The deposit products include 25K money market accounts (MM) and 10K CD accounts with 3-month, 6-month and 12-month maturity (3M CD, 6M CD, 12M CD). The underlying data are from Ratewatch, Call Reports, HMDA and FRED. The sample period is from 2001Q1 to 2019Q4. Fixed effects are denoted at the bottom of the table. Standard errors clustered by county are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.11: Robustness checks for deposit rate regressions: Large banks

Dependent Variable:	Δbranch deposit rate (quarterly)							
Loan market power:	Bank-HMDA-HHI							
Deposit product:	MM		3M CD		6M CD		12M CD	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1: $HHI \times \Delta i$	-0.715	0.589	1.499*	3.527***	1.636*	1.497	3.093***	3.239***
	(0.755)	(1.392)	(0.794)	(1.358)	(0.863)	(1.260)	(0.836)	(1.130)
β_2: $HHI \times i \times \Delta i$	0.484*	0.390	-0.504	-0.589*	-0.453	-0.453	-0.927***	-0.926***
	(0.292)	(0.302)	(0.321)	(0.331)	(0.340)	(0.338)	(0.309)	(0.306)
β_3: $HHI \times E \times \Delta i$		-9.712		-16.735*		2.668		0.119
		(9.449)		(9.544)		(8.789)		(7.206)
Obs	93,397	93,220	92,442	92,442	97,496	97,496	97,752	97,752
Adj R^2	0.986	0.987	0.970	0.970	0.962	0.962	0.961	0.961
Branch FE	Y	Y	Y	Y	Y	Y	Y	Y
Bank FE	Y	Y	Y	Y	Y	Y	Y	Y
County×time FE	Y	Y	Y	Y	Y	Y	Y	Y
Bank controls	Y	Y	Y	Y	Y	Y	Y	Y

Notes: This table presents the estimation of equation (1.44) for branch deposit rates using the data of the largest 25% banks by inflation-adjusted quarterly average assets. In the estimation the interaction terms with bank equity-assets ratio are included. The dependent variable is the quarterly change of a branch's deposit rate on a deposit product. For the independent variables listed in the table, HHI is the bank-level average Herfindahl index of home mortgage loans across counties (Bank-HMDA-HHI). i is the nominal interest rate. E is the bank equity-assets ratio. Δi is the quarterly change in the Fed funds rate. The sample consists of all U.S. counties with branches of at least two different banks for identification. The deposit products include 25K money market accounts (MM) and 10K CD accounts with 3-month, 6-month and 12-month maturity (3M CD, 6M CD, 12M CD). The underlying data are from Ratewatch, Call Reports, HMDA and FRED. The sample period is from 2001Q1 to 2019Q4. Fixed effects are denoted at the bottom of the table. Standard errors clustered by county are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.12: Net effects of the channel: branch deposit rates

Dependent Variable:	Δbranch deposit rate (quarterly)			
Loan market power:	Bank-HMDA-HHI			
Deposit product:	MM	3M CD	6M CD	12M CD
	(1)	(2)	(3)	(4)
$HHI \times \Delta i$	0.541***	2.330***	2.637***	2.681***
	(0.185)	(0.244)	(0.222)	(0.227)
$HHI \times i \times \Delta i$	-0.071	-0.702***	-0.789***	-0.855***
	(0.074)	(0.092)	(0.083)	(0.084)
$i \times \Delta i$	0.061***	0.149***	0.170***	0.183***
	(0.005)	(0.006)	(0.006)	(0.006)
Obs	251,477	240,852	259,087	259,793
Adj R^2	0.988	0.973	0.965	0.963
Controls (all panels):				
Branch FE	Y	Y	Y	Y
Bank FE	Y	Y	Y	Y
County FE	Y	Y	Y	Y
Bank controls	Y	Y	Y	Y

Notes: This table presents the estimated net effects of loan market concentration and nominal interest rate on deposit rate passthrough using equation (1.44). The dependent variable is the quarterly change of a branch's deposit rate on a deposit product. For the independent variables listed in the table, HHI is the bank-level average Herfindahl index of home mortgage loans across counties (Bank-HMDA-HHI). i is the nominal interest rate. E is the bank equity-assets ratio. Δi is the quarterly change in the Fed funds rate. The sample consists of all U.S. counties with branches of at least two different banks for identification. The deposit products include 25K money market accounts (MM) and 10K CD accounts with 3-month, 6-month and 12-month maturity (3M CD, 6M CD, 12M CD). The underlying data are from Ratewatch, Call Reports, HMDA and FRED. The sample period is from 2001Q1 to 2019Q4. Fixed effects are denoted at the bottom of the table. Standard errors clustered by county are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.13: Identification of the channel: branch deposits growth

Dependent Variable	$\Delta \log$ (Branch Deposits, Annual)							
Loan market power	Bank-HMDA-HHI							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1: $HHI \times \Delta i$	-0.043*	0.073	0.266***	0.263***	0.268***	0.148**	0.176***	0.108***
	(0.023)	(0.064)	(0.085)	(0.086)	(0.068)	(0.059)	(0.059)	(0.042)
β_2: $HHI \times i \times \Delta i$		-0.059**	-0.059**	-0.054**	-0.065***	-0.045***	-0.037**	-0.038**
		(0.025)	(0.024)	(0.025)	(0.018)	(0.017)	(0.017)	(0.017)
β_3: $HHI \times E \times \Delta i$			-1.924***	-1.965***	-2.015***	-1.200**	-1.458***	
			(0.611)	(0.614)	(0.541)	(0.473)	(0.479)	
$i \times \Delta i$								0.006***
								(0.001)
Obs	1,108,167	1,108,167	1,108,167	1,108,167	1,108,167	1,108,167	1,108,167	1,108,167
Adj R^2	0.190	0.190	0.191	0.188	0.157	0.147	0.032	0.147
Controls (all panels):								
Branch FE	Y	Y	Y	Y	Y	Y	N	Y
Bank FE	Y	Y	Y	Y	Y	Y	N	Y
County	N	N	N	N	Y	Y	Y	Y
Time FE	N	N	N	N	N	Y	Y	N
County×time FE	Y	Y	Y	Y	N	N	N	N
State×time FE	N	N	N	N	Y	N	N	N
Bank controls	Y	Y	Y	N	Y	Y	Y	Y

Notes: This table presents the estimation of equation (1.44) for branch deposit growth. The dependent variable is the annual log difference of a branch's total deposit volumes. For the independent variables listed in the table, HHI is the bank-level average Herfindahl index of home mortgage loans across counties (Bank-HMDA-HHI). i is the nominal interest rate. E is the bank equity-assets ratio. Δi is the quarterly change in the Fed funds rate. The sample consists of all U.S. counties with branches of at least two different banks for identification. Deposit growth is equal to the symmetric growth rate of branch deposits at annual frequency. The underlying data are from FDIC, Call Reports, HMDA and FRED. The sample period is from 1997 to 2019. Fixed effects are denoted at the bottom of the table. Standard errors clustered by county are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.14: Robustness checks of regressions on deposits growth

Dependent Variable	$\Delta \log$ (Branch Deposits, Annual)							
Loan market power	Bank-HMDA-HHI							
Robustness test	Nominal Rate: FF		Nominal Rate: T-Bill		Pre-financial crisis		Large banks	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1: $HHI \times \Delta i$	0.032	0.191***	0.057	0.206***	0.147	0.233	0.106	0.453***
	(0.038)	(0.065)	(0.042)	(0.067)	(0.180)	(0.207)	(0.083)	(0.114)
β_2: $HHI \times i \times \Delta i$	-0.052***	-0.050***	-0.057***	-0.054***	-0.095*	-0.086	-0.071**	-0.080**
	(0.018)	(0.018)	(0.019)	(0.019)	(0.057)	(0.057)	(0.032)	(0.032)
β_3: $HHI \times E \times \Delta i$		-1.605***		-1.515***		-1.119		-3.339***
		(0.559)		(0.552)		(0.871)		(0.828)
Obs	1,108,167	1,108,167	1,108,167	1,108,167	381,383	381,383	958,410	958,410
Adj R^2	0.198	0.198	0.198	0.198	0.322	0.323	0.203	0.204
Controls (all panels):								
Branch FE	Y	Y	Y	Y	Y	Y	Y	Y
Bank FE	Y	Y	Y	Y	Y	Y	Y	Y
County×time FE	Y	Y	Y	Y	Y	Y	Y	Y
Bank controls	Y	Y	Y	Y	Y	Y	Y	Y

Notes: This table presents the robustness tests on the branch deposit growth regression (1.44). In Column (1) and (2), the nominal interest rate is the yearly Fed funds rate. In Column (3) and (4), the nominal interest rate is the 1-year Treasury yield rate. In Column (5) and (6), the estimation uses data before the 2008 global financial crisis (until June 2008). In Column (7) and (8), the estimation uses data of largest 25% banks by inflation-adjusted annual average assets. The dependent variable is the annual log difference of a branch's total deposit volumes. For the independent variables listed in the table, HHI is the bank-level average Herfindahl index of home mortgage loans across counties (Bank-HMDA-HHI). i is the nominal interest rate. E is the bank equity-assets ratio. Δi is the quarterly change in the Fed funds rate. The sample consists of all U.S. counties with branches of at least two different banks for identification. Deposit growth is equal to the symmetric growth rate of branch deposits at annual frequency. The underlying data are from FDIC, Call Reports, HMDA and FRED. The sample period is from 1997 to 2019. Fixed effects are denoted at the bottom of the table. Standard errors clustered by county are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.15: Determinants of the threshold nominal interest rate

Dependent Variable:	Δbranch deposit rate (quarterly)			
Deposit product:	MM	3M CD	6M CD	12M CD
Regressor x of $HHI \times x \times \Delta i$	(1)	(2)	(3)	(4)
Equity/assets	-20.282***	-13.426**	-3.949	-6.379*
	(7.077)	(6.438)	(4.921)	(3.844)
Bank size (log assets)	0.274***	0.345***	0.079	0.398***
	(0.086)	(0.094)	(0.100)	(0.085)
Loan/assets	0.067	-0.609	-1.098	0.965
	(1.222)	(1.310)	(1.354)	(1.200)
Core deposits/liabilities	4.769**	-2.215	-5.174**	-3.961**
	(2.271)	(2.001)	(2.082)	(1.819)
Maturity gap	-0.054	-0.102	-0.047	0.086
	(0.084)	(0.089)	(0.079)	(0.076)
Nonperforming loan share	18.133**	17.135	20.769**	24.602***
	(8.576)	(10.684)	(8.295)	(7.652)
Other assets	-0.284	2.132	-0.579	1.174
	(1.748)	(2.061)	(2.225)	(1.845)
Other liabilities	0.672	-0.729	-1.203	-2.348***
	(0.892)	(1.067)	(0.934)	(0.854)
Obs	204,674	194,101	212,148	212,824
Adj R^2	0.988	0.974	0.966	0.966
All FEs	Y	Y	Y	Y
Bank controls	Y	Y	Y	Y

Notes: This table reports the effects of bank characteristics on the threshold value of nominal interest rate, below which the passthrough monetary policy rate to deposit rate increases in a bank's loan market concentration. The vector of coefficients of $HHI \times x \times \Delta i$ are reported in the table, where x represents the following bank characteristics: equity-assets ratio, log of total assets, loan-assets ratio, core deposits share in total liabilities, maturity gap, the share of nonperforming loans in total loans, the share of other assets in total interest-earning assets, the share of other liabilities in total liabilities. The underlying data are from HMDA, Ratewatch, FRED and Call Reports. The sample period is 2001Q1 to 2019Q4. Standard errors clustered by county are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.16: Regression on the state-dependent passthrough

Balance sheet component	Deposits	Assets	Securities	Loans	RE loans	C&I loans
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Zero beta, all banks						
Deposit rate	0.059***	0.049***	0.096***	0.026***	0.018	0.025**
	(0.008)	(0.007)	(0.015)	(0.009)	(0.020)	(0.011)
Obsevations	4,885	4,885	4,885	4,885	4,885	4,885
R^2	0.015	0.013	0.010	0.002	0.000	0.002
Panel B: Slope beta, all banks						
Deposit rate	0.089***	0.069***	0.125***	0.047***	0.059***	0.054***
	(0.005)	(0.005)	(0.011)	(0.006)	(0.014)	(0.007)
Observations	4,885	4,885	4,885	4,885	4,885	4,885
R^2	0.072	0.055	0.033	0.018	0.004	0.015
Panel C: Zero beta, large banks						
Deposit rate	0.105***	0.076***	0.056*	0.087***	0.116***	0.098***
	(0.022)	(0.020)	(0.034)	(0.021)	(0.031)	(0.026)
Obsevations	244	244	244	244	244	244
R^2	0.088	0.061	0.012	0.066	0.056	0.063
Panel D: Slope beta, large banks						
Deposit rate	0.106***	0.081***	0.088***	0.086***	0.120***	0.100***
	(0.021)	(0.019)	(0.033)	(0.019)	(0.032)	(0.024)
Obsevations	244	244	244	244	244	244
R^2	0.100	0.079	0.031	0.078	0.057	0.077

Notes: This table reports the relationship between the state-dependent passthrough of monetary policy rates to bank deposit rates and the state-dependent transmission of monetary policy rates to bank balance sheet components. The analysis covers all U.S. commercial banks in the Call Reports data with at least 60 quarters observations from 1997Q1 to 2019Q4, excluding the periods of global financial crisis (2008Q3 to 2009Q2). Panel A and B report the resutsl for all banks, Panel C and D report the results for the largest 5% of banks by assets. The passthrough of monetary policy rates to bank deposit rates is measured as the deposit rate zero beta and slope beta. The zero beta measures the sensitivity of a bank's deposit rate to the Fed funds rate when the Fed funds rate is zero. The slope beta measures the change in the sensitivity when the Fed funds rate increases by 100 bps. The state-dependent transmission of monetary policy rates to bank balance sheet components include the analogous zero betas and slope betas for the quarterly growth of a bank's deposits (column (1)), assets (column (2)), securities (column (3)), loans (column (4)), real estate loans (column (5)) and commercial and industrial loans (column (6)). Panel A and C report the coefficient of regressing a balance sheet component's zero beta on deposit rate zero beta. Panel B and D report the coefficient of regressing a balance sheet component's slope beta on deposit rate slope beta. All the betas are winsorized at 5%. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

CHAPTER 2

Disintermediating the Federal Funds Market

with Tsz-Nga Wong

2.1 Introduction

This paper endogenizes the search intensity in the random matching model of Federal funds market in Afonso and Lagos (2015b). This model features a novel *disintermediation* channel: when the unconventional monetary policy like the interest of excess reserves (IOER) or the balance sheet cost like the liquidity requirement of Basel III or the FDIC assessment fee increases, banks cease to intermediating Federal funds for the efficient use. This disintermediation channel is missing in the random matching model of Afonso and Lagos (2015b). In particular, the latter predicts changes in the IOER and balance sheet cost have no effect, and the equilibrium is always constrained efficient. In a general environment of costly search, we establish the analytical properties like supermodularity and constrained inefficiency. In a particular environment, we solve both the equilibrium and constrained optimum in closed forms, which allow us to characterize the comparative dynamics on the trading volume, Federal fund rates, and the reserves distribution, as well as to identify the sources of inefficiency.

Our model is motivated by the following empirical evidence. A Federal funds trade is intermediated if the purchasing bank is also selling Federal funds on the same day, i.e., the bank borrows reserves from one bank to lend reserves to other banks. Since the forth quarter of 2008, the trade volume in the Federal funds market has been shrinking sharply, which is

largely driven by the decline in intermediated trades, as illustrated in Figure 2.3. The decline happened in both the level of intermediated trades and the number of intermediating banks. We refer this disappearance of intermediated trades as the *disintermediation* of the Federal funds market. We notice that the timing of the *disintermediation* coincides with a series of unconventional monetary policies, which start with the introduction of IOER, followed by three rounds of quantitative easing (QE) as well as the changes in regulation, such as the introduction of Basel III and widening of the basis of the FDIC's deposit insurance assessment fee. The latter increases the balance sheet cost of holding reserves. To identify the effects of these policies on the Federal funds trades and intermediated trades, we perform a series of instrumental variable regressions on a panel dataset of bank-level Federal funds trade volume. The dataset is collected from various sources, such as FFIEC Call Reports, Form FR-Y9C and SEC 10-Qs and 10-Ks. We find that the unconventional monetary policies significantly lower the level of intermediated trades on both extensive and intensive margin, and also impede the allocation of Federal funds from net lenders to net borrowers. These findings are robust to alternative specifications.

However, according to the standard random matching model, any changes in the interest on excess reserves or balance sheet cost of holding reserves do not affect the level of intermediation – all the effects are absorbed in the changes of the Federal funds rates. Furthermore, the random matching model predicts that the vast increase of reserves injected by QEs should have increased the level of intermediation instead. The reason is that, since matching is costless in Afonso and Lagos (2015b), banks always search for counterparties in the market, and they always trade to split their reserve holdings equally once they match with each other. Therefore, the level of trades, along both the extensive and intensive margins, does not change even though the introduction of IOER or balance sheet cost changes the marginal value of holding reserves, as long as it is diminishing. It also implies that banks should trade more reserves when their average holding of reserves increases proportionally, ceteris paribus. Cost-free search means in the constrained efficient allocation, banks should

always search and share the reserve holdings equally, coinciding with the equilibrium allocation. We show that these features no longer hold when putting search intensity becomes costly.

While the disintermediation effect of transaction costs may seem straightforward, the disintermediation effect of the unconventional monetary policy calls for an explanation. It is puzzling since it is commonly thought that government sponsored enterprises (GSE) like Feddie Mae and Federal Home Loan Banks are not entitled to the IOER. It implies that there should be more Federal funds trades between GSEs and non-GSEs, and intermediated loans in general, to earn the abitrage of the IOER. Our theory is that unconventional monetary policy also amplifies the disintermediation effect of transaction costs. To illustrate this, we build a continuous-time costly search and bargaining model of the over-the-counter unsecured loan market. The baseline model admits a closed-form solution, which allows for sharp comparative statics. In this case, the IOER reduces the volume of intermediated loans, raises the average level of the Fed fund rates but reduces their dispersion. Balance sheet cost and regulation cost reduce the volume but have ambiguous effects on the level and dispersion of the Fed fund rates. Also, with costly endogenous search, theoretically there could be multiple equilibria; in particular, no trading is always an equilibrium. We propose a refinement that always selects the most "liquid" equilibrium and prove its existence and uniqueness in the general model.

We further calibrate our theoretical model with the empirical data via simulated method of moments, and conduct counterfactual analysis to evaluate the magnitudes of unconventional monetary policies and regulations on the disintermediation. We find that the disintermediation is mostly driven by IOER and the rising transaction cost, while the effect of excess reserve balances is small. In particular, in the year of 2018, the share of intermediation volume in total Federal funds volume doubles if we decrease IOER to its 2006 level (which is zero), and the share increases by four times if we decrease the estimated transaction costs to the 2006 level.

Literature. Our paper relates to several strands of literature. First, starting with Poole (1968), there has been a series of researches on the Federal funds market in partial equilibrium or general equilibrium models. Hamilton (1996) provides a partial equilibrium model to study the effects of transaction costs on the daily dynamics of the Federal funds rates. More recently, some studies focus on the monetary policy implementation and passthrough efficiency in the environment of excess reserves, such as Duffie and Krishnamurthy (2016), Bech and Keister (2017). In the meantime, other papers discuss the role of interbank markets and unconventional monetary policies on the aggregate outcome and welfare, such as Kashyap and Stein (2012), Ennis (2018), Williamson (2019), Bigio and Sannikov (2021) and Bianchi and Bigio (forthcoming).

Another strand of literature focuses on capturing the over-the-counter (OTC) nature of the Federal funds markets and its implications. On the one hand, some reserches develop two-sided matching models to capture the search and matching frictions between lenders and borrowers, such as Berentsen and Monnet (2008), Bech and Monnet (2016), Afonso et al. (2019) and Chiu et al. (2020). These models are able to fit a number of aggregate empirical moments of the interbank markets in the U.S. and Europe and provide fruitful policy implications. However, the intermediation trades, which are important features of OTC markets, are missing in those models. On the other hand, people use continuous-time one-sided matching models to capture the intermediation feature of OTC markets. The one-sided matching models are pioneered by the seminal works of Afonso and Lagos (2015b) and Afonso and Lagos (2015a). Our model endogenizes the time-varying search intensity to study the disintermediation trades. The related papers include Duffie et al. (2005), Lagos and Rocheteau (2009), Trejos and Wright (2016), Farboodi et al. (2017), Lagos and Zhang (2019), Üslü (2019), Hugonnier et al. (2020) and Liu (2020).

There have been other papers that use network approach to study the interbank markets. For example, Bech and Atalay (2010) explores the network topology of the Federal funds market, and Gofman (2017) builds a network-based model of the interbank lending market

and quantifies the efficiency-stability trade-offs of regulating large banks. Chang and Zhang (2018) develops a dynamic model that allows agents to endogenously choose counterparties and form network structure. They find that some agents specialize in market making and become the core of the financial network, with the purpose of eliminating information frictions.

Outline. The remainder of the paper is as follows. Section 2.2 describes the institutional background and the aggregate empirical facts that motivate our paper. Section 2.3 documents the empirical evidence on the disintermediation effect of unconventional monetary policies at the individual bank level. Section 2.4 presents the theoretical framework for our analysis. Section 2.5 provides a class of models that allows for closed-form solutions and comparative statics. Section 2.6 structurally estimates the analytical model and quantitatively decomposes the effects of unconventional monetary policies on Federal funds intermediation. The final section, 2.7, concludes the paper.

2.2 The landscape of the Federal funds market

This section introduces the institutional features, the policy and regulatory environment and the aggregate trade dynamics in the Federal funds market to motivate our estimation and theoretical model in the following sections. We will focus on the change of the landscape of this market before and after the Great Recession as the market has changed drastically since then. To measure the aggregate and composition of the Fed funds trade activity, we aggregate the data from a set of regulatory filings, including the quarterly Consolidated Report of Condition and Income for U.S. banks and branches (Call reports), the Consolidated Financial Statements (Form FR Y-9C) for bank holding companies (BHC) and SEC 10-Ks and 10-Qs for other eligible entities.

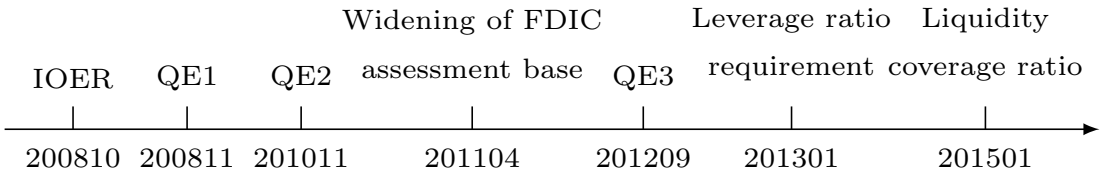
2.2.1 Institutional background

The Federal funds market is a market for unsecured loans of dollar reserves held at the Federal Reserve Banks. The market interest rates on these loans are commonly referred to as the Federal funds rates. Most of the Federal funds transactions are overnight (99%). Financial institutions (FIs) rely on the Fed funds market for short-term liquidity needs: First, the Federal funds is not considered as the deposits to the borrower bank under Regulation D, thus it is useful for borrower banks to satisfy their reserve requirements and payments needs. Second, the lender FIs can lend excess reserves and earn overnight Fed funds rate. Regarding the market structure, the Federal funds market is an over-the-counter (OTC) market without centralized exchange. A borrower bank (Federal funds purchased) and a lender bank (Federal funds sold) meet and trade bilaterally, and the transfer of funds is completed through the Fed's reserve accounts.

The market of Federal Funds has been the epicenter where monetary policies are implemented. Before the Great Recession, the Federal Reserve adjusted the supply of reserve balances, by the purchase and sale of securities in the open market, so as to keep the Fed fund rates around the target of monetary policy. Since the Great Recession, the landscape of the market has changed drastically due to a series of unconventional monetary policies and regulations. Figure 2.1 plots the timeline of these changes, which start with the introductin of interest on excess reserves (IOER), followed by three rounds of quantitative easing (QE) as well as changes in regulation, such as the widening of the basis for FDIC assessment fee and the introduction of Basel III regulations.

Due to the changes in policy and regulations, the Fed funds market has entered a stage with excess reserves, and the Federal Reserve relies on two new policy tools to implement its desired target range for the Federal funds rate: the IOER, which it offers to eligible depository institutions, is set at the top of the target ranges; and the rate of return at the overnight reverse repurchase (ON RRP) facility, which is available to an expanded set of counterparties

Figure 2.1: Timeline of unconventional monetary policy and regulation

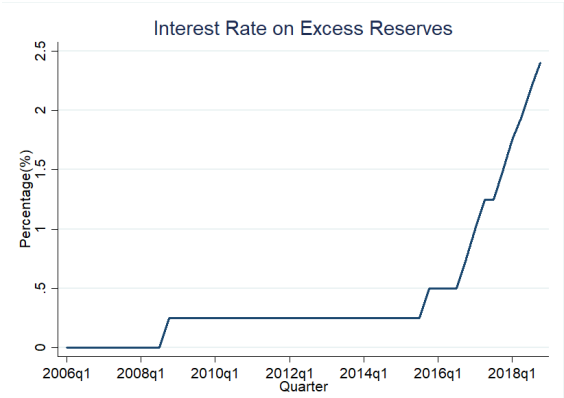


Notes: This figure plots the timeline of unconventional monetary policies and regulations since the Great Recession. The numbers on the timeline represents the date (year-month) when the policy or regulation is introduced.

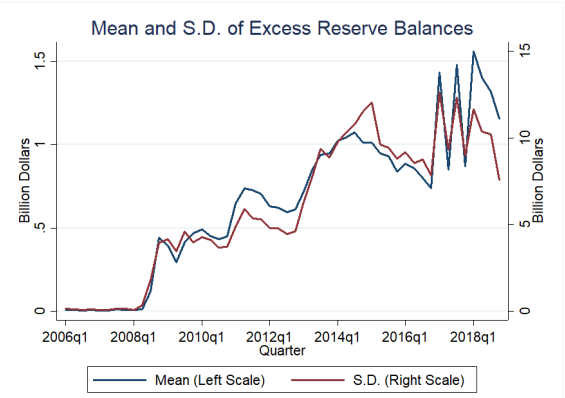
including government-sponsored enterprises (GSEs) and some money market funds, is set at the bottom of the rage. Figure 2.2 shows the time series of the unconventional monetary policies. Panel (a) plots the path of IOER, which has been steadily increasing between 2008Q4 and 2018Q4. Panel (b) plots the mean and standard deviation of individual excess reserve balances in the same period , which has grown drastically since the Great Recession.

Figure 2.2: IOER and Excess Reserves

(a) IOER



(b) Excess reserves



Notes: This figure plots the sequences of IOER, the mean and standard deviation of individual excess reserve balances from 2006Q1 to 2018Q4. Data source: FRED, Call reports, FR Y-9C.

2.2.2 (Dis)intermediation in the Federal funds market

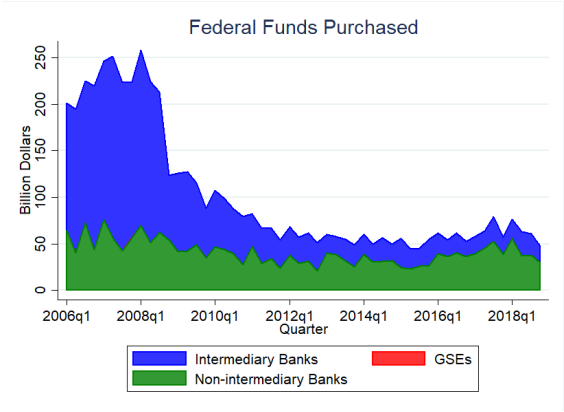
Due to the over-the-counter structure, the Fed funds trades involve a significant share of intermediation trading. A group of banks act as intermediaries by borrowing reserves from the lender banks and lending them to others on the same day. We find that the intermediary banks are responsible for most of the decline in Fed funds volume. Specifically, by consolidating the individual balance sheet data, we decompose the total Fed funds volume into three groups: intermediary banks, non-intermediary banks and government-sponsored enterprises (GSEs). As illustrated in Panel (a) of Figure 2.3, the decline of Fed funds purchased (borrowing) is entirely driven by intermediary banks, whose volume of borrowing sharply declined from the peak of $195 billion in 2007Q2 to an average of $22 billion in 2018. At the same time, the volume of borrowing by other groups stayed stable over time. Panel (b) of Figure 2.3 suggests that, on the supply side, the depository institutions account for most of the decline of Fed funds lending. In particular, the lendings by intermediary banks was more than $60 billion on average before 2008, but decreased sharply to almost zero right after the Great Recession. The non-intermediary banks accounted for about $50 billion lending before the Great Recession, and shrank gradually to less than $5 billion over time in 2018.

The decline of borrowing and lending by intermediaries imply the decline of Fed funds reallocation. We find that this decline occurs on both extensive and intensive margin. As plotted in Panel (a) of Figure 2.4, a significant share (more than 15%) of Fed funds volume is traded for intermediating purposes. However, since the financial crisis has declined by more than two thirds to less than 5%. Moreover, Panel (b) of Figure 2.4 shows that the number of intermediary banks has also decreased from 600 in 2006 to less than 100 at the end of 2018.

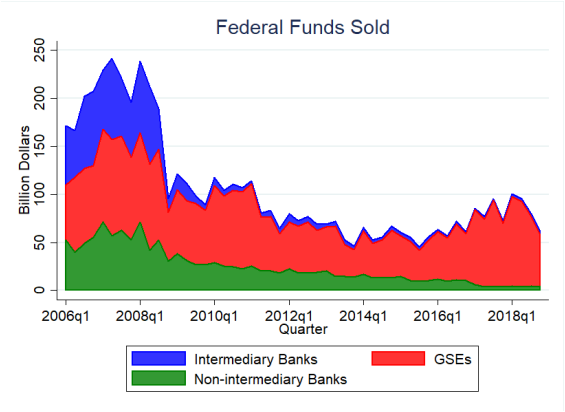
Why did disintermediation happen? Certainly, the Federal funds market has been going through a transition from the Great Recession, but it is worth noting that the timing of the disintermediation coincides with the changes in the monetary policies and regulations, as plotted in Figure 2.1. All these changes closely relate to banks' incentive to trade Fed

Figure 2.3: Decomposition of Federal funds volume

(a) Federal funds purchased



(b) Federal funds sold



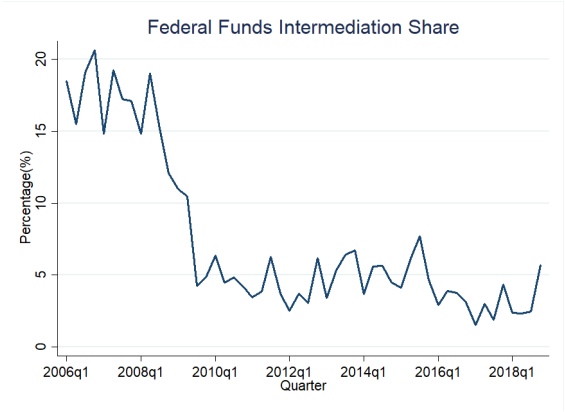
Notes: This figure plots the decomposition of the aggregate Federal funds purchased and sold by groups from 2006Q1 to 2018Q4. Data source: Call reports, FR Y-9C, SEC 10-K and 10-Q.

funds. For example, the introduction of IOER raises the return of holding reserves, which lowers banks' lending incentives and raises their borrowing incentives. The QEs have left banks flush with excess reserves. As a result, the demand for borrowing reserves to meet the reserve requirement and payment needs has become rare. The regulation changes The widening of FDIC assessment base and Basel III regulations increase the balance sheet cost of holding reserves. For example, FDIC insurance premium is now charged according to the size of FI's assets (instead of the size of deposit), which is increasing in the Federal Funds borrowed. Furthermore, Basel III now imposes a cap on the FI's leverage ratio and a floor of the holding of liquid (and usually low-return) asset to cover potential cash outflow, increasing the regulation cost.

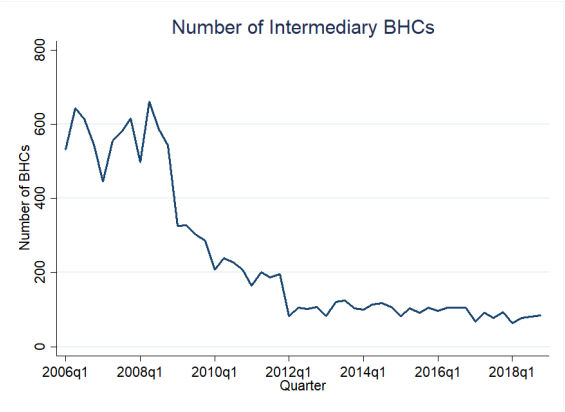
Our empirical facts about the disintermediation coincide with the existing literature. For example, Keating and Macchiavelli (2017) find that the proportion of intermediated funds declined sharply after the financial crisis. On the daily level, the domestic banks keep more than 99% percent of Fed funds borrowed and foreign banks keep more than 80%. These evidence document the importance of intermediation to the substantial decline of Fed funds

Figure 2.4: Aggregate Fed funds intermediation

(a) Intermediation volume share



(b) Number of intermediary banks



Notes: This figure plots the intermediation volume share and number of intermediary banks. Data source: Call reports, FR Y-9C.

volume. We will focus on examining banks' incentive to intermediate and its implications for the monetary policy implementation.

2.3 Empirical evidence

In this section, we document the empirical relationship of Federal funds intermediation trades and the unconventional monetary policies. Our focus is to test the following hypotheses using U.S. bank-level data described in Section 2.3.1.1:

Hypothesis 1 The number of intermediary banks and the individual bank's volume of Federal funds intermediation decrease in IOER and the aggregate excess reserves.

Based on the facts shown in Figure 2.4, the first hypothesis tests the causal effect of IOER and the aggregate excess reserves on the intermediation trading in Federal funds market. We examine the impact on both extensive margin and intensive margin. In addition to testing the impact on intermediation trades, we also investigate whether the disintermediation effect

of IOER and aggregate excess reserves affect the allocation of reserves between the reserve net lenders and net borrowers.

Hypothesis 2 A higher IOER and aggregate excess reserves lower the net Federal funds purchased by net borrowers (banks that have net borrowing of Federal funds) and the net Federal funds lent by net lenders (banks that have net lending of Federal funds).

This hypothesis examines whether borrower banks are less able to find lenders if the intermediation trades decrease. The following sections describe the data and estimation results.

2.3.1 Data

2.3.1.1 Bank-level data

The bank-level financial data are collected from various sources. We use the quarterly Consolidated Report of Condition and Income for U.S. banks and branches (commonly known as "Call reports") and Consolidated Fiancial Statemennts (Form FR Y-9C) for Bank Holding Companies (BHCs).[1] The call reports and Form FR Y-9C are quarterly filed with the Federal Reserve by all U.S. banks and branches, and form FR Y-9C is filed by all U.S. holding companies with total consolidated assets of $1 billion or more (prior to 2015, this threshold was $500 million. Since September 2018, this number changes to $3 billion). These files report the balance sheet data of US banks at the end of each quarter, including the Federal funds purchased (Fed funds borrowing), Federal funds sold (Fed funds lending) and other balance sheet characteristics. Given the Fed funds are mostly overnight, the volume of Fed funds trade reported in these files measures banks' Fed funds borrowing and lending on the last business day of each quarter. Our data covers the period going from 2003Q1 to

[1] Appendix 2.A describes the detailed data source and construction process.

2018Q4.[2] We measure each variable at the consolidated top holder level. Aggregating the variables to the top holder level not only avoids double counting, but also eliminates the bilateral trades between subsidiaries of a bank holding company that are not implemented in the Fed funds market.

For each top holder in each quarter, we construct the following variables: (1) Net volume of Fed funds purchased normalized by total assets ($ffnet\_assets$), i.e.

$$ffnet\_assets = \frac{\text{Fed funds purchased} - \text{Fed funds sold}}{\text{total assets}}.$$

It measures a bank's net borrowing of Fed funds as a share of bank assets. (2) Volume of Fed funds reallocation normalized by total assets ($ffreallo\_assets$), i.e.

$$ffreallo\_assets = \frac{\text{Fed funds purchased} + \text{Fed funds sold}}{\text{total assets}} - |ffnet\_assets|.$$

This variable follows the definition of Fed funds reallocation in Afonso and Lagos (2015b), which is equal to the Fed funds trade in excess of the net borrowing. (3) Excess reserve balances before Federal funds trade normalized by total assets before Fed funds trade, i.e.

$$exres\_assets = \frac{\text{excess reserve balances before Federal funds trade}}{\text{total assets}}.$$

The excess reserve balances before Federal funds trade represent a bank's holdings of Federal reserves balances in excess of its reserve requirement when it enters the Federal funds market. It captures individual heterogeneity of trade incentives in the Fed funds market. It is equal to a bank's excess reserve balances recorded in the bank balance sheets minus the net Federal funds purchased (Federal federal funds purchased minus Federal funds sold). Moreover, for individual controls, we include the following balance-sheet variables: (1) logged value of total assets ($\log\_assets$); (2) total loans normalized by total assets ($loan\_assets$); (3) total

[2]We also use the data in 2002Q4 as the lagged values of variables in 2003Q1.

nonperforming loans normalized by total assets (npl_assets); (4) total high-quality liquid assets normalized by total assets ($hqla\_assets$); (5) total equity normalized by total assets ($equi\_assets$); (6) tier-1 leverage ratio ($tier1\_lev\_ratio$); (7) ROA (roa); (8) dummies of top holders' entity types ($entity\_type$).

2.3.1.2 Aggregate-level data

We use two sets of aggregate variables. The first set includes Interest Rate on Reserves ($ioer$), Primary Credit Rate(dw), quarterly real GDP growth rate ($rgdpg$), quarterly unemployment rate ($unemp$), standard deviation of the Fed's general treasury account in a quarter. All these variables are measured at the end of a quarter. The interest rate on excess reserves and primary credit rate are the main regressors of monetary policy. They represent the outside return of holding reserves by lender banks and borrower banks at the end of a trading session, respectively. The other variables are the aggregate controls in regressions.

The second set of aggregate-level variables are obtained from bank-level data. For the cross section of top holders in each quarter, we construct the moments of excess reserve distribution: (1) aggregate excess reserves normalized by aggregate bank assets (agg_exres_assets); (2) standard deviation of excess reserve balances normalized by the mean (sd_exres_norm =S.D. of excess reserves/Mean of excess reserves);[3] (3) skewness of excess reserve distribution (sk_exres). The aggregate excess reserves agg_exres_assets is the third main regressor of monetary policy. It captures the effect of the Fed's total reserve balances on Fed funds trade. Meanwhile, we control the standard deviation and skewness to capture the effect of reserve distribution.

[3]Using standard deviation of excess reserves normalized by average assets produces similar results.

2.3.2 Effects on intermediation trade

Our first specification explores the impact of IOER and aggregate reserves on banks' intermediation trading. Note that in the data sample, only a fraction of banks are intermediaries, and the measure of individual bank's intermediation, $ffreallo\_assets$, is non-negative. Thus we study how IOER and aggregate reserves impact both the probability of intermediation trades (extensive margin) and the volume of intermediation (intensive margin). In particular, we run probit and tobit regressions on the following specification on the sample of banks that hold positive total reserves at the Fed account and intermediate Federal funds at least once in the data sample:

$$\begin{aligned} y_{i,t} &= fixed\_effects + \beta_0 exres\_assets_{i,t} + \beta_1 ioer_t + \beta_2 ioer_t \times exres\_assets_{i,t} \\ &\quad + \beta_3 agg\_exres\_assets_t + \beta_4 agg\_exres\_assets_t \times exres\_assets_{i,t} \\ &\quad + \beta_5 dw_t + \beta_6 dw_t \times exres\_assets_{i,t} + \gamma \cdot controls_{i,t} + \varepsilon_{i,t}, \end{aligned} \tag{2.1}$$

where $y_{i,t} = \mathbf{1}\{ffreallo\_assets_{i,t} > 0\}$ in probit regression, and $y_{i,t} = ffreallo\_assets_{i,t}$ in tobit regressions. The term $fixed\_effects$ represents the fixed effects on bank entity type, Fed district, and the time dummies for 2008 financial crisis and post-crisis periods. By adding the interaction between the policy variables and individual excess reserve balances, we also investigate the potential heterogeneous effects of the unconventional monetary policies across banks.

The probit and tobit estimation assumes exogeneity of the regressors. However, the Fed funds trade volume could depend on unobserved factors that correlate with the main regressors. For example, a bank's Federal funds trade volume and excess reserve balances could be driven by some common unobserved factors, e.g. sophistication of balance sheet management. Moreover, a bank's incentive to trade Federal funds could be driven by some unobserved aggregate shocks that are correlated with the changes in IOER, primary credit rate and aggregate excess reserves. Thus we augment the estimation with instrumental-

variable probit and tobit regressions to examine the potential endogeneity of excess reserves, aggregate policies and Federal funds trades. First, the instruments for IOER and primary credit rate are the cumulative monetary policy shocks (policy news shocks and Federal funds rate shocks) over past 4 quarters, which are obtained from Nakamura and Steinsson (2018).[4] Second, the instrument for the aggregate excess reserves is the one-period lag of 4-quarter change in aggregate excess reserves to aggregate bank assets ratio. Third, the instrument for individual excess reserves is one-period lag of individual excess reserves. For the instruments of interaction terms, we use the interactions between the corresponding instruments mentioned above.

The results of probit regressions are shown in Table 2.3, where we report three groups of estmation: column (1) and (2) reports the standard Probit estimation, Column (3) and (4) report the estimation of a random effects panel probit model, and column (5) and (6) report the estimation of the instrumental-variable probit model. In all columns, the probability of intermediation trade decreases in IOER and the aggregate excess reserves. The coefficients are significant and robust. The primary credit rate also negatively impacts the probability of intermediation trade, and the coefficient is significant in the random effect estimator and IV tobit estimator. This implies that the unconventional monetary policies have strong disintermediation effect on the intensive margin. Moreover, by adding the interaction terms, we find that the impact of IOER and aggregate excess reserves on the probability of intermediation trade can be heterogeneous across banks, but the signs of the coefficients for the interaction terms are not consistent and robust across the columns.

The results of tobit regressions are reported in Table 2.4, where we also have three groups of estimation. The main results of tobit regressions are similar to those of probit regressions. On average, under a higher value of IOER, aggregate excess reserves and primary credit rate, banks are less likely to do intermediation tradeS. The coefficients are significantly negative

[4]The original sample period of the policy shocks end in 2014, and Acosta and Saia (2020) update the shocks to 2019. We use the later in our estimation.

and robust across columns. In summary, the estimation results of probit and tobit regressions imply significantly and consistently negative effect of unconventional monetary policies on intermediation trade, which reveals a strong disintermediation channel.

2.3.3 Effects on net borrowing of Federal Funds

Our second specification relates the net Fed funds borrowing to a bank's excess reserve balances, IOER and aggregate reserve balances. We estimate the following equation on the sample of banks that hold positive total reserves at the Fed account and trade Federal funds at least once in the data sample:

$$\begin{aligned} ffnet\_assets_{i,t} &= \alpha_i + \eta_{yr(t)} + \beta_0 exres\_assets_{i,t} + \beta_1 ioer_t + \beta_2 ioer_t \times exres\_assets_{i,t} \\ &\quad + \beta_3 agg\_exres\_assets_t + \beta_4 agg\_exres\_assets_t \times exres\_assets_{i,t} \\ &\quad + \beta_5 dw_t + \beta_6 dw_t \times exres\_assets_{i,t} + \gamma \cdot controls_{i,t} + \varepsilon_{i,t}, \end{aligned} \tag{2.2}$$

where i represents a bank and t denotes the last business day of a quarter. The parameters α_i and $\eta_{yr(t)}$ represent the bank fixed effects and year fixed effects. The control variables $controls_{i,t}$ include both the bank-level controls and the aggregate controls mentioned above. This regression examines how the level of IOR and aggregate excess reserves impact individual banks' net Fed funds borrowing. By adding the interaction between the policy variables and individual excess reserve balances, we also investigate the potential heterogeneous effects of the monetary policies across banks.

Colums (1) to (3) of Table 2.5 report the results of OLS estimation. Column (1) does not include the interaction terms, thus estimates the average effect of the monetary policies on banks' net Fed funds borrowing. Column (2) reports the estimation of our baseline specification (2.2), while Column (3) additionally controls the quarter fixed effects. We have the following findings. First, the coefficient of individual excess reserves, β_0, is significantly negative across all the columns. It implies that banks with more excess reserves borrow

less Fed funds. Second, the OLS estimation shows significant and robust heterogeneous effects of monetary policies on net Fed funds borrowing. In particular, the coefficients of the interaction between IOR and individual excess reserves, β_2, and the interaction between the aggregate excess reserves and individual excess reserves, β_4, are both positive. Moreover, the coefficient of the interaction between primary credit rate and individual excess reserves, β_6, is negative. It means that for banks with sufficiently high reserve balances, their net borrowing increases in IOR and the aggregate excess reserves, and decreases in primary credit rate. On the other hand, for banks with sufficiently low reserve balances, their net borrowing decreases in IOR and aggregate excess reserves, and increases in primary credit rate. Since banks with high (low) excess reserves are more likely to be net Fed funds lenders (borrowers), the estimation results imply that a higher IOER and aggregate excess reserves impede the reallocation of Fed funds from lender banks to borrower banks. On the other hand, a higher primary credit rate enhances the reallocation of Fed funds.

Column (4) to (6) of Table 2.5 report the results of 2SLS estimation, where the specification of each column corresponds to Column (1) to (3). The results are consistent with the OLS estimation. In Column (4), we find that banks net Fed funds borrowing decreases in IOER, primary credit rate and aggregate excess reserves on average. In Column (5) and (6), the coefficients of all interaction terms are significant and consistent with the OLS estimation. Thus our estimation documents robust negative effect of IOER and aggregate excess reserves as well as positive effect of primary credit rate on Federal funds allocation. This verifies our second hypothesis.

2.4 A search model of Federal funds market

Overview. In this section we propose a theoretical framework for our analysis. The timing and preferences of the framework follow Afonso and Lagos (2015b), but we endogenize the

banks' search intensity.[5] A Federal funds market runs continuously from time 0 to T. A unit-measure of banks starts the Federal funds market with idiosyncratic level of reserve balances, $k_0 \in \mathbb{K} = [k_{\min}, k_{\max}] \subset \mathbb{R}$, following a cummulative distribution F_0. There is also a numéraire good, where banks can consume and produce linearly at time $T + \Delta$. Why do banks trade reserve balances? Holding reserve balances k_t at t yields a flow payoff $u(k_t)$ continuously from time 0 to T, and also a terminal payoff $U(k_T)$ at time T, which is affected by (unconventional) monetary policy and reserve requirement, as we will see in the next section. Thus, banks with a higher maringal value of reserves want to purchase reserves balances (Federal funds) and settle in numéraire later at time $T + \Delta$.[6] However, trading in the Federal funds market is subject to search frictions. In particular, it takes time for a bank to find but a random counterparty such that the evolution of reserve balances follows a jump process:

$$k_t = k_0 + \sum_{t_n \leq t} q_{t_n}, \tag{2.3}$$

where t_n is the Poisson time of finding the n-th counterparty, from whom the bank purchases q_{t_n} (sells if negative) units of reserves balances. As we will see, the search friction is essential to generate the dispersion of Federal funds rates, slow trades, and intermediation we observe in practice.

Search. Time-varying contact rate is an important feature of the Federal funds market. Before the Great Recession, most of the Federal funds trades happened in the late afternoon, which suggests that search intensity is higher when t is close to T. Time-varying search intensity also suggests that Federal funds market could be vulnerable to gridlock, which is captured by the search externality of the matching function.

In the model, a pair of banks is matched at the Poisson arrival rate $m(\varepsilon_t, \varepsilon_t')$ at t, where

[5] We also allow reserve balances being divisible rather than discrete.

[6] Following the terminology in Call Reports, we use the terms Federal funds purchased (sold) and reserve balances borrowed (lent) interchangeably.

ε_t and ε'_t are their search intensities.[7] We normalize that $\varepsilon \in [0,1]$ with $m(0,0) = \lambda_0$, $m(1,0) = \lambda_1$, and $m(1,1) = \lambda$. We assume that the matching function is symmetric, increasing, supermodular, and additive in counterparty's search intensity such that

$$m\left[\varepsilon, \alpha\varepsilon' + (1-\alpha)\varepsilon''\right] = \alpha m(\varepsilon, \varepsilon') + (1-\alpha) m(\varepsilon, \varepsilon''). \tag{2.4}$$

Define the search profile of all k-banks as $\varepsilon_t = \{\varepsilon_t(k)\}_{k\in\mathbb{K}}$. By additivity, a bank with seach intensity ε_t matches *some* counterparties at the rate $m(\varepsilon_t, \bar{\varepsilon}_t)$, where $\bar{\varepsilon}_t \equiv \int \varepsilon_t(k')\, dF_t(k')$ is the average search intensity of banks at t. It captures the search complementarity effect.

Our leading examples are $m(\varepsilon, \varepsilon') = \lambda_0 + (\lambda - \lambda_0)(\varepsilon + \varepsilon')/2$ and $m(\varepsilon, \varepsilon') = \lambda_0 + (\lambda - \lambda_0)\varepsilon\varepsilon'$.[8] Some matches are "free", which arrive at the rate λ_0. Both examples capture the fact that a bank can search for a bank or be found by others. The former assumes that the likelihoods of finding a bank and being found are independent, each proportional to the bank's and the counterparty's search intensity, respectively. The latter assumes that the likelihoods of finding a bank and being found are the same, which are proportional to both the bank's and the counterparty's search intensity.

---

[7] For readers not familar with the Poisson model, the probabality that a bank exerting a contingent plan of search intensity $\{\varepsilon_t\}_{t=0}^T$ until its next trade will find a counterparty bank within τ units of time is

$$\Pr\{t_1 \leq \tau\} = 1 - \exp\left\{-\int_0^\tau \int_{j\in[0,1]} m\left(\varepsilon_t, \varepsilon_t^j\right) djdt\right\}.$$

[8] The general form of the matching function is

$$m(\varepsilon, \varepsilon') = (\lambda - 2\lambda_1 + \lambda_0)\varepsilon\varepsilon' + (\lambda_1 - \lambda_0)(\varepsilon + \varepsilon') + \lambda_0.$$

See Appendix 2.C.1 for derivations.

Preferences. The individual bank's problem is given by

$$\max_{\varepsilon} \mathbb{E}^{\varepsilon} \left\{ \int_0^T e^{-rt} u(k_t)\, dt + e^{-rT} U(k_T) - \sum_{n=1,2,\ldots} \left[e^{-rt_n} \chi(\varepsilon_{t_n}, q_{t_n}) + e^{-r(T+\Delta)} R_{t_n} \right] \right\}, \text{ s.t. (2.3).} \tag{2.5}$$

The terms in the brackets of (2.5) are the expected discounted payoff flow from holding reserves, the discounted terminal payoff of holding reserves at time T, the discounted cost of trading q_{t_n} units of Federal funds at search intensity ε_{t_n} with the n-th counterparty at t_n, and the repayment R_{t_n} in numéraire to settle these trades. The dynamics of reserve balances (k_t) is given by (2.3). The amount of Federal funds traded and its repayment are determined by Nash bargaining protocol when the bank finds its n-th counterparty at t_n. The bank's problem is to choose a contingent plan of search intensity (ε) to maximize the expected discounted payoff (2.5).

The payoff functions, u and U, are positive, continuously differentiable, increasing, concave and at least one of them is strictly concave. The cost function $\chi(\varepsilon, q)$ is positive, continuously differentiable in both arguments, convex in q, complementary in ε and q, and satisfies Inada condition in q. We normalize that $\chi(0, q) = \chi(\varepsilon, 0) = 0$, and assume symmetry over q, i.e. $\chi(\varepsilon, q) = \chi(\varepsilon, -q)$. Note that Afonso and Lagos (2015b) is the special case of $\chi(\varepsilon, q) = 0$. The cost function captures the fact that it is increasingly costly to trade fast and large in the Federal funds market. Notice that the cost is incurred when the match and trade happen. As we will see later, this feature generates a tension between cost shifting and seach complementarity.

Bargaining. Once a bank meets a counterparty, the terms of trade (q_t, R_t) are negotiated according to the Nash bargaining protocol. Denote $V_t(k)$ as the maximal attainable contin-

uation value of a bank holding k units of reserve balances at t.[9] For this bank, the trade surplus of purchasing q units of reserve balances with R units of numéraire repayment from its counterparty at t is

$$B_t(k,q,R,\varepsilon) \equiv V_t(k+q) - e^{-r(T-t+\Delta)}R - \chi(\varepsilon,q) - V_t(k).$$

By symmetry, dentoe $B_t(k',-q,-R,\varepsilon')$ as the trade surplus of its counterparty whose reserves balance before trade is k'. The terms of trade solve the following Nash bargaining problem:

$$\max_{\substack{q,R\in\mathbb{R}\\ k+q,k'-q\in\mathbb{K}}} B_t(k,q,R,\varepsilon)\, B_t(k',-q,-R,\varepsilon'). \tag{2.6}$$

Denote the solution as $q_t = q_t(k,k',\varepsilon,\varepsilon')$ and $R_t = R_t(k,k',\varepsilon,\varepsilon')$. Thus, for all $k \in \mathbb{K}$ and $t \in [0,T]$, the value function is given by

$$V_t(k) = \mathbb{E}^{\varepsilon}\left\{\begin{array}{l} \int_0^{\min\{t_{+1},T\}-t} e^{-r\tau}u(k)\,d\tau + 1_{t_{+1}>T}e^{-r(T-t)}U(k) \\ +1_{t_{+1}\le T}e^{-r(t_{+1}-t)}\int \left\{\begin{array}{l} V_{t_{+1}}\left[k+q_{t_{+1}}\left[k,k',\varepsilon_{t_{+1}},\varepsilon_{t_{+1}}(k')\right]\right] \\ -\chi\left[\varepsilon_{t_{+1}},q_{t_{+1}}\left[k,k',\varepsilon_{t_{+1}},\varepsilon_{t_{+1}}(k')\right]\right] \\ -e^{-r(T+\Delta-t_{+1})}R_{t_{+1}}\left[k,k',\varepsilon_{t_{+1}},\varepsilon_{t_{+1}}(k')\right] \end{array}\right\} \\ \times\frac{m\left[\varepsilon_{t_{+1}},\varepsilon_{t_{+1}}(k')\right]}{m\left(\varepsilon_{t_{+1}},\bar{\varepsilon}_{t_{+1}}\right)}dF_{t_{+1}}(k') \end{array}\right\}, \tag{2.7}$$

[9] While the terminology is standard, to be precise, the value function is defined as

$$V_t(k) \equiv e^{rt}\max_{\varepsilon}\mathbb{E}_t^{\varepsilon}\left\{\int_t^T e^{-rz}u(k_z)\,dz + e^{-rT}U(k_T) - \sum_{t_n\ge t}\left[e^{-rt_n}\chi(\varepsilon_{t_n},q_{t_n}) + e^{-r(T+\Delta)}R_{t_n}\right]\right\} \text{ given } k_t = k.$$

where

$$\begin{aligned} q_t(k,k',\varepsilon,\varepsilon') &= \arg\max_q \{V_t(k+q)+V_t(k'-q)-\chi(\varepsilon,q)-\chi(\varepsilon',q)\}, \\ e^{-r(T+\Delta-t)}R_t(k,k',\varepsilon,\varepsilon') &= \frac{1}{2}\left\{\begin{array}{l} V_t[k+q_t(k,k',\varepsilon,\varepsilon')]-V_t(k)-\chi[\varepsilon,q_t(k,k',\varepsilon,\varepsilon')] \\ V_t(k')-V_t[k'-q_t(k,k',\varepsilon,\varepsilon')]+\chi[\varepsilon',-q_t(k,k',\varepsilon,\varepsilon')] \end{array}\right\}, \end{aligned}$$

and t_{+1} is the random time of matching the next counterparty, arriving at the rate $m(\varepsilon_t,\bar{\varepsilon}_t)$. The costs of search intensities, $\chi(\varepsilon,q)$ and $\chi(\varepsilon',q)$, are shared in the bargaining; it creates the cost shifting effect.

Define the Federal funds rate as $\rho_t(k,k',\varepsilon,\varepsilon') \equiv R_t(k,k',\varepsilon,\varepsilon')/q_t(k,k',\varepsilon,\varepsilon')-1$. Note that the bargaining solution is symmetric, i.e.,

$$q_t(k,k',\varepsilon,\varepsilon') = -q_t(k',k,\varepsilon,\varepsilon') = -q_t(k',k,\varepsilon',\varepsilon) = q_t(k',k,\varepsilon',\varepsilon),$$

and $\rho_t(k,k',\varepsilon,\varepsilon') = \rho_t(k',k,\varepsilon',\varepsilon)$. Denote the joint surplus as

$$\begin{aligned} S_t(k,k',\varepsilon,\varepsilon') &\equiv V_t[k+q_t(k,k',\varepsilon,\varepsilon')]-V_t(k)-\chi[\varepsilon,q_t(k,k',\varepsilon,\varepsilon')] \\ &\quad +V_t[k'-q_t(k,k',\varepsilon,\varepsilon')]-V_t(k')-\chi[\varepsilon',-q_t(k,k',\varepsilon,\varepsilon')]. \end{aligned}$$

Due to the linear preferences in R, banks split the joint surplus evenly such that

$$\begin{aligned} & B_t[k,q_t(k,k',\varepsilon,\varepsilon'),R_t(k,k',\varepsilon,\varepsilon'),\varepsilon] = B_t[k',-q_t(k,k',\varepsilon,\varepsilon'),-R_t(k,k',\varepsilon,\varepsilon'),\varepsilon'] \\ = \; & 0.5S_t(k,k',\varepsilon,\varepsilon'). \end{aligned}$$

Given the assumption on the cost function χ, the following lemma characterizes the property of the bargaining solution.[10]

[10]The proofs of all the propositions and lemmas are provided in Appendix 2.C.

Lemma 2.1 *(i). $S_t(k,k',\varepsilon,\varepsilon')$ and $|q_t(k,k',\varepsilon,\varepsilon')|$ are both decreasing in ε and ε'. Moreover, suppose $V_t(k)$ is weakly concave and twice differentiable, then $S_t(k,k',\varepsilon,\varepsilon')$ is supermodular in ε and ε'.*

(ii). If $V_t(k)$ is (strictly) concave, then $S_t(k,k,\varepsilon,\varepsilon')=0$, and $S_t(k,k',\varepsilon,\varepsilon')$ is (strictly) decreasing in k for all $k<k'$ and (strictly) increasing in k for all $k>k'$. We have $q_t(k,k',\varepsilon,\varepsilon')>0$ and is decreasing in k and increasing in k'.

Value and distribution. Given the search profile of banks and the trade surplus function, the value function, $V_t(k)$, of (2.7) can be recursively expressed as the solution the following Hamiltonian-Jacob-Bellman (HJB) equation[11]

$$rV_t(k)=\dot{V}_t(k)+u(k)+\max_{\varepsilon_t\in[0,1]}\int\frac{1}{2}S_t\left[k,k',\varepsilon_t,\varepsilon_t(k')\right]m\left[\varepsilon_t,\varepsilon_t(k')\right]dF_t(k'), \tag{2.8}$$

where $V_T(k)=U(k)$. The initial value $V_0(k_0)$ equals (2.5).

Given the search profile and the bargaining solution, by counting the inflow and outflow, the balance distribution satisfies the following Kolmogorov forward equation (KFE)[12]

$$\dot{F}_t(k^w)=\left\{\begin{array}{c}\int_{k>k^w}\int m\left[\varepsilon_t(k),\varepsilon_t(k')\right]1\left\{k+q_t(k,k')\leq k^w\right\}dF_t(k')\,dF_t(k)\\ -\int_{k\leq k^w}\int m\left[\varepsilon_t(k),\varepsilon_t(k')\right]1\left\{k+q_t(k,k')>k^w\right\}dF_t(k')\,dF_t(k)\end{array}\right\}, \tag{2.9}$$

given $F_0(k^w)$. The intuition of the KFE is as follows. Consider two groups of banks: those holding not greater than k^w units of reserve balances $I_-(k^w)$ and the rest $I_+(k^w)$, so the measure of $I_-(k^w)$ at t is $F_t(k^w)$. The first line of (2.9) is the inflow rate to $I_-(k^w)$ post-trade from $I_+(k^w)$ pre-trade; the second line of (2.9) is the outflow rate from $I_-(k^w)$ pre-trade to

[11]For readers not familiar with the HJB equation, we derive (2.8) in the online Appendix 2.C.3. The discretized version of (2.8) without search cost or transaction cost is Proposition 1 of Afonso and Lagos (2015b).

[12]For readers not familiar with the KFE, we derive (2.9) in the online Appendix 2.C.3. When k is discrete, $F_t(k)$ is probability mass function shown in Proposition 2 of Afonso and Lagos (2015b).

$I_+(k^w)$ post-trade.

2.4.1 Equilibria

The terms of trade and choices of search intensity interact with the dynamics of reserve distribution in the Federal funds market. The feedback mechanism is summarized by the system of forward-looking value functions, V_t, and backward-looking distribution functions, F_t. We define a symmetric subgame perfect equilibrium as follows.

Definition 2.1 *An equilibrium consists of* $\{V_t(k), \varepsilon_t(k), F_t(k), q_t(k,k'), \rho_t(k,k')\}_{k,k' \in \mathbb{K}, t \in [0,T]}$ *such that,*

(a) given $\{\varepsilon_t(k'), F_t(k'), q_t(k,k'), \rho_t(k,k')\}_{k,k' \in \mathbb{K}, t \in [0,T]}$*, the value function* $V_t(k)$ *solves the bank's maximization problem (2.8) with* $\varepsilon_t = \varepsilon_t(k)$ *at all* t*;*

(b) given $\{V_t(k)\}_{k \in \mathbb{K}, t \in [0,T]}$*,* $q_t(k,k')$ *and* $\rho_t(k,k')$ *solve the Nash bargaining problem (2.6);*

(c) given $\{\varepsilon_t(k), q_t(k,k')\}_{k,k' \in \mathbb{K}, t \in [0,T]}$*, the distribution function* $F_t(k)$ *satisfies (2.9).*

Multiplicity. Even the equilibrium exists, yet to prove, there are multiple equilibria for, at least, three reasons. First, due to the dynamic complementarity, it is well-known that a system of forward-backward differential equations can have multiple solutions.[13] Second, due to the search complementarity (m is supermodular), the higher search intensities put by other banks the higher marginal propensity to match. Third, due to the cost shifting (S

[13]For example, consider a simple system of forward-backward ODEs:

$$\begin{aligned} \dot{y}(t) &= -x(t), \text{ where } y(2\pi) = 0, \\ \dot{x}(t) &= y(t), \text{ where } x(0) = 0, \end{aligned}$$

which has a continuum of solutions $\{x(t) = A\sin t, y(t) = A\cos t\}$. In macroeconomics, the literature of equilibrium indeterminacy after the seminar work of Benhabib and Farmer (1994) has illustrated various possibilities of multiplicity in standard neo-classical growth models consisting of, typically, a system of forward-looking (the capital accumulation) and backward-looking differential equations (the Euler equation). Here our economy deals with a more complex system of partial different equations: the state variable is the distribution of reserves, instead of capital, thus the dimension is infinite, instead of one.

is supermodular), the higher search cost shared by other banks the lower the marginal cost of search intensity, as less Federal funds are traded. To see it, using (2.8), the equilibrium search profile is a fixed point function to the following functional:

$$\Gamma_t(\varepsilon_t)(k) \equiv \arg\max_{\varepsilon\in\{0,1\}} \left\{ \int S_t[k,k',\varepsilon,\varepsilon_t(k')]\, m[\varepsilon,\varepsilon_t(k')]\, dF_t(k') \right\}. \tag{2.10}$$

Denote the set of fixed points to Γ_t as $\Omega(S_t,F_t) \subseteq [0,1]^{\mathbb{K}}$, i.e., $\varepsilon_t(k) = \Gamma_t(\varepsilon_t)(k)$ for all $\varepsilon_t \in \Omega(S_t,F_t)$. To proceed we need some notions of lattice theory. Consider two search profiles $\varepsilon(k)$ and $\varepsilon'(k)$. Define a partial order $\succeq_s$ such that $\varepsilon \succeq_s \varepsilon'$ if $\varepsilon(k) > \varepsilon'(k)$ for all $k \in \mathbb{K}$. A lattice $\{\mathbb{L}, \succeq_s\}$ is complete if for any $\varepsilon, \varepsilon' \in \mathbb{L} \subseteq [0,1]^{\mathbb{K}}$, we have either $\varepsilon \succeq'_s \varepsilon$ or $\varepsilon' \succeq_s \varepsilon$. Suppose S_t satisfies the conditions for supermodularity in $(\varepsilon,\varepsilon')$ as in Lemma 2.1. The following proposition provides a sufficiently condition for complete lattice.

Proposition 2.1 *$\Omega(S_t,F_t)$ is non-empty. Suppose the cost function $\chi(\varepsilon,q)$ is separable, i.e. $\chi(\varepsilon,q) = \kappa(\varepsilon)\tilde{\chi}(q)$. Define $\theta_\kappa(\varepsilon) \equiv \kappa'(\varepsilon)\varepsilon/\kappa(\varepsilon)$, $\theta_m(\varepsilon) \equiv m_{12}(\varepsilon,\varepsilon')\varepsilon/m_2(\varepsilon,\varepsilon')$ and $X(k,k',\varepsilon,\varepsilon') \equiv S_t(k,k',\varepsilon,\varepsilon') / \{(\kappa(\varepsilon)+\kappa(\varepsilon'))\tilde{\chi}[q_t(k,k',\varepsilon,\varepsilon')]\}$.*

Given S_t and F_t,$\{\Omega(S_t,F_t), \succeq_s\}$ is a complete lattice if $X(k,k',\varepsilon,\varepsilon') \geq 1$ and

$$\theta_\kappa(\varepsilon) \leq \theta_m(\varepsilon).$$

In other words, for any two equilibrium search profiles ε and ε', they can always be ranked by Proposition 2.1 such that it is either $\varepsilon_t(k) \geq \varepsilon'_t(k)$ for *all* k, or $\varepsilon_t(k) \leq \varepsilon'_t(k)$ for *all* k. Define the largest equilibrium profile of $\Omega(S_t,F_t)$ as $\varepsilon_t(k)$ if $\varepsilon_t(k) \succeq_s \varepsilon'_t(k)$ for all $\varepsilon'_t(k) \in \Omega(S_t,F_t)$. Notice that no-search equilibrium exists even when there is no search cost ($\chi = 0$), as a result of coordination failure. Although the cardinality of $\Omega(S_t,F_t)$ is potentially large, the supermodularity of the search game implies a lattice structure to classifiy the equilibria for analysis. To deal with multiplicity, most of time we will focus on the following equilibrium refinement.

Definition 2.2 *An equilibrium satisfies the defreezing refinement if there is no other equilibrium with a strictly higher average search intensity of banks. If $\Omega(S_t, F_t)$ is a complete lattice, then an equilibrium satisfies the defreezing refinement is also the largest equilibrium profile of $\Omega(S_t, F_t)$ for all $t \in [0, T]$.*

The defreezing refinement addresses the multiplicity due to matching complementarity. No-search equilibrium is always eliminated by the defreezing refinement if other equilibria exist. Although the uniqueness of the equilibrium is not guaranteed under the defreezing refinement due to the forward-backward differential equation system, we show in Section 2.5 that under the defreezing refinement, we are able to obtain a class of models with closed-form solutions. The closed-form model allows for comparative statics that are consistent with the empirical evidence.

2.4.2 Efficiency

The equilibrium is not necessarily efficient, even the one that satisfies the defreezing refinement. Consider a social planner that dictates search decision $\{\varepsilon_t^p(k)\}$ and bilateral exchange of reserve balances $\{q_t^p(k, k')\}$ to maximize the discounted sum of the utility flows of banks with equal weights, taking as given the search frictions and transaction costs.

Definition 2.3 *A constrained efficient allocation consists of $\{\varepsilon_t^p(k), F_t^p(k), q_t^p(k, k')\}_{k,k' \in \mathbb{K}, t \in [0,T]}$ that solves*

$$\mathbb{W} = \max \left\{ \begin{array}{l} \int_0^T e^{-rt} \int u(k)\, dF_t^p(k)\, dt + e^{-rT} \int U(k)\, dF_T^p(k) \\ -\int_0^T \int \int e^{-rt} \chi\left[\varepsilon_t^p(k), q_t^p(k, k')\right] m\left[\varepsilon_t^p(k), \varepsilon_t^p(k')\right] dF_t^p(k')\, dF_t^p(k)\, dt \end{array} \right\} \quad (2.11)$$

subject to the law of motion of reserves

$$\dot{F}_t^p(k^w) = \left\{ \begin{array}{c} \int_{k>k^w} \int m\left[\varepsilon_t^p(k), \varepsilon_t^p(k')\right] 1\left\{k + q_t^p(k, k') \leq k^w\right\} dF_t^p(k')\, dF_t^p(k) \\ -\int_{k \leq k^w} \int m\left[\varepsilon_t^p(k), \varepsilon_t^p(k')\right] 1\left\{k + q_t^p(k, k') > k^w\right\} dF_t^p(k')\, dF_t^p(k) \end{array} \right\}, \quad (2.12)$$

where $F_0^p(k^w) = F_0(k^w)$.

The constrained efficient allocation $\{\varepsilon_t^p(k), q_t^p(k,k')\}$ maximizes the Hamiltonian. Denote $V_t^p(k)$ as the co-state to $dF_t^p(k)$, the Hamiltonian is given by

$$\begin{aligned}\mathcal{H}_t^p \equiv & \int u(k)\, dF_t^p(k) - \int\int \chi\left[\varepsilon_t^p(k), q_t^p(k,k')\right] m\left[\varepsilon_t^p(k), \varepsilon_t^p(k')\right] dF_t^p(k')\, dF_t^p(k) \\ & + \int\int m\left[\varepsilon_t^p(k), \varepsilon_t^p(k')\right] \left\{V_t^p\left[k + q_t^p(k,k')\right] - V_t^p(k)\right\} dF_t^p(k')\, dF_t^p(k) \\ & + \int\int \eta_t(k,k') \left[q_t^p(k,k') - q_t^p(k',k)\right] dF_t^p(k')\, dF_t^p(k),\end{aligned} \tag{2.13}$$

where $\eta_t(k,k')$ is the multiplier to the bilateral trade constraint $q_t^p(k,k') + q_t^p(k',k) = 0$. The evolution of the co-state solves[14]

$$\begin{aligned}rV_t^p(k) = & \dot{V}_t^p(k) + u(k) \\ & + \int \left\{ \begin{array}{l} V_t^p\left[k + q_t^p(k,k')\right] + V_t^p\left[k' - q_t^p(k,k')\right] - V_t^p(k) \\ -V_t^p(k') - \chi\left[\varepsilon_t^p(k), q_t^p(k,k')\right] - \chi\left[\varepsilon_t^p(k'), -q_t^p(k,k')\right] \end{array} \right\} \\ & \times m\left[\varepsilon_t^p(k), \varepsilon_t^p(k')\right] dF_t^p(k'),\end{aligned} \tag{2.14}$$

with $V_T^p(k) = U(k)$. The optimal allocation $\{q_t^p(k,k')\}_{k,k' \in \mathbb{K}}$ satisfies

$$q_t^p(k,k') = \arg\max_q \left\{V_t^p(k+q) + V_t^p(k'-q) - \chi\left(\varepsilon_t^p(k), q\right) - \chi\left(\varepsilon_t^p(k'), -q\right)\right\}$$

and the optimal search profile $\{\varepsilon_t^p(k)\}_{k \in \mathbb{K}}$ is a fixed point function to

$$\varepsilon_t^p(k) = \Gamma_t^p\left(\varepsilon_t^p\right)(k) \equiv \arg\max_{\varepsilon \in [0,1]} \left\{\int S_t^p\left(k, k', \varepsilon, \varepsilon_t^p(k')\right) m\left(\varepsilon, \varepsilon_t^p(k')\right) dF_t^p(k')\right\},$$

[14] We derive (2.14) in the Appendix 2.C.5.

where

$$S_t^p(k,k',\varepsilon,\varepsilon') = \max_q \left\{ \begin{array}{c} V_t^p(k+q) - V_t^p(k) - \chi(\varepsilon,q) \\ +V_t^p(k'-q) - V_t^p(k') - \chi(\varepsilon',-q) \end{array} \right\}.$$

Note that the equilibrium HJB (2.8) for $V_t(k)$ differs from the co-state HJB (2.14) since the gains from bilateral trade in the co-state HJB is double of that in the equilibrium HJB. The following proposition shows that in general the equilibrium allocation is not constrained optimal – the welfare theorem is violated.

Proposition 2.2 ***(Inefficiency)*** *Equilibrium is not generically constrained optimal. Equilibrium is constrained optimal if* $\chi = 0$.

Afonso and Lagos (2015b) show that the welfare theorem holds when banks are homogeneous (beyond initial balances); Proposition 2.2 shows it is no longer the case when there is search cost or transaction cost. Üslü (2019) shows that the welfare theorem does not hold when banks are ex-ante heterogeneous in, for example, payoff functions and contact rates, because of the composition externality. Proposition 2.2 shows that even banks are ex-ante homogeneous, the welfare theorem still does not hold when banks can choose their contact rates or when Federal funds trades are subject to transaction cost.

2.4.3 Walrasian benckmark

To see the role of search intensity, consider the Walrasian benchmark where there is no search fricton ($\lambda_0 = \infty$) and trades are organized in a competitive market. Banks are free to trade at any $t \in [0,T]$, taking the competitive Federal funds rates ρ_t^w as given. It will be useful to express the bank's problem in term of its value of reserve balances, $a_t \equiv (1+\rho_t^w)\, k_t$. The evolution of a_t is thus given by

$$da_t = \frac{\dot{\rho}_t^w}{1+\rho_t^w} a_t dt + d\delta_t, \tag{2.15}$$

where the first term is the appreciation of the reserve value due to the appreciation of Federal funds rate and the second term, δ_t, is the value of Federal fund purchased up to t. Notice that we allow $d\delta_t$ to be infinitesimal or lumpy. At $T+\Delta$ the bank will settle the accumulated Federal funds purchased, which is δ_T. Similar to (2.5), given the path of competitive Federal funds rates $\{\rho_t^w\}$, the bank problem is given by

$$\max_{\{\delta_t\}} \mathbb{E}\left\{\int_0^T e^{-rt} u\left(\frac{a_t}{1+\rho_t^w}\right) dt + e^{-rT} U\left(\frac{a_T}{1+\rho_T^w}\right) - e^{-r(T+\Delta)}\delta_T\right\}, \text{ s.t. (2.15).} \tag{2.16}$$

Denote $\delta_t(a_0)$ as the solution chosen at t by a bank that holds a_0 units of reserve value at $t=0$. In the competitive equilibrium, ρ_t^w clears the market clearing such that for all t

$$0 = \int \delta_t\left[\left(1+\rho_0^w\right)k\right] dF_0(k). \tag{2.17}$$

Proposition 2.3 *In the competitive equilibrium, we have*

(a) $\rho_t^w = e^{r\Delta}\left\{U'(K) + \left[e^{r(T-t)}-1\right]\frac{u'(K)}{r}\right\} - 1$;

(b) $\delta_t(a) = \left(1+\rho_0^w\right)K - a$ *for all* $t \in [0,T]$.

In the Walrasian benchmark, banks trade instantaneously at $t=0$ such that every bank maintains K units of reserve balances throughtout the horizon. In the competitive equilibrium, the Federal funds rate is decreasing over time, in order to compensate for the utility from holding reserve. Also, notice that the Walrasian benchmark is the first-best allocation.

2.5 A class of closed-form models

This section develops a closed-form model based on the theoretical framework. This model provides comparative statics that are consistent with our empirical evidence.

Preferences, monetary policy and regulation. We assume quadratic payoff functions, which are given by

$$\begin{aligned} u(k) &= -a_2 k^2 + a_1 k, \\ U(k) &= -A_2 k^2 + A_1 k. \end{aligned}$$

The matching function is given by

$$m(\varepsilon, \varepsilon') = (\lambda - \lambda_0)\, \varepsilon \varepsilon' + \lambda_0.$$

The cost function is given by

$$\chi(\varepsilon, q) = \kappa \varepsilon q^2.$$

The parameter κ captures various balance sheet costs of purchasing Federal funds in practice. For instance, κ captures the regulatory cost of purchasing Federal funds by reducing the leverage ratio and liquidity coverage ratio, as required by the Basel III regulation. Moreover, Dodd-Frank act mandates FDIC to widen the assessment base of its deposit insurance premium to bank's consolidated total assets (previously, the assessment base consisted of the domestic deposit only). For the reserves lenders ($q < 0$), selling Federal funds will not the size of their total assets (substituting the liability of Federal Reserve Banks with the liability of other banks). For the reserve borrowers ($q > 0$), purchasing Federal funds increases the size of their total assets (in term of reserves balances) so they pay additional deposit insurance premium.

Unconventional monetary policy in practice consists of paying IOER and central bank liquidity facility like primary dealer credit and, traditionally, discount window. Basel III regulation also encourages the holdings of HQLA like reserves. To model these, we assume that there are $k_+ K, k_- K \in \mathbb{K}$ such that $U'(k_+ K) = 1 + i^{ER} + \gamma$ and $U'(k_- K) = 1 + i^{DW} + \gamma$, where i^{ER} the interest rate on excess reserve and, i^{DW}, where $i^{DW} > i^{ER}$, is the interest

rate of the liquidity facility, and γ is the regulatory benefit of holding reserve balances. In practice, k_+K is the level of reserves sufficiently excess the reserve requirement to collect the IOER; k_-, where $k_- < k_+$, is the level of reserves sufficiently below the reserve requirement such that the bank is penalized by, for example, the discount window rate. The simplest differentiable specification capturing the above is given by

$$A_2 \equiv \frac{i^{DW} - i^{ER}}{2K\left(k_+ - k_-\right)}, A_1 \equiv 1 + \frac{k_+ i^{DW} - k_- i^{ER}}{k_+ - k_-} + \gamma.$$

Under the above specification, we first guess (and verify later) that the value function admits a closed-form solution, which is quadratic in k but with time-varying coefficients.

Bargaining solution. Given a quadratic value function, the bargaining solution is given by

$$q_t\left(k, k', \varepsilon, \varepsilon'\right) = \left\{1 - \underbrace{\left[1 - \frac{V_t''(k) + V_t''(k')}{2\kappa\left(\varepsilon + \varepsilon'\right)}\right]^{-1}}_{\text{precaution-speed trade-off}}\right\} \underbrace{\frac{k' - k}{2}}_{\text{efficient bilateral trade}}, \tag{2.18}$$

$$1 + \rho_t\left(k, k', \varepsilon, \varepsilon'\right) = \underbrace{e^{r(T+\Delta-t)}}_{\text{time cost}} \left[\underbrace{\frac{V_t'(k) + V_t'(k')}{2}}_{\text{sharing marginal valuation}} + \underbrace{\kappa \frac{\varepsilon' - \varepsilon}{2} q_t\left(k, k', \varepsilon, \varepsilon'\right)}_{\text{speed premium (discount)}}\right] \tag{2.19}$$

In Afonso and Lagos (2015b), the meeting banks exchange the efficient trade size $\frac{k'-k}{2}$ and leave with the same post-trade reserve balances. Moreover, due to the equal bargaining power, they trade at the price equal to the average of their marginal valuations of reserves. However, in the existence of transaction cost and endogenous search intensity, the bilateral trade size is less than the efficient level. The trade size is decreasing in κ and the meeting banks' search intensity ε and ε', since a higher κ, ε and ε' imply higher marginal cost of transaction. The effect of search intensity on trade size captures the precaution-speed trade-

off. With a higher search intensity, banks are able to find counterparties faster and also more costly. Thus they respond by covering orders with smaller size in each transaction.

At the same time, the endogenous search intensity also induces a speed premium or discount of the bilateral Fed funds rate, which is similar to Üslü (2019). The premium is proportional to the trade size and the difference in the search intensities between the counterparties. In the meetings with $k' > k$ and $\varepsilon' > \varepsilon$, or $k' < k$ and $\varepsilon' < \varepsilon$, the seller bank searches faster than the buyer bank. This generates a positive speed externality for the buyer while the seller pays a higher cost. The Nash bargaining creates a cost shifting from seller to buyer and the bilateral Fed funds rate is charged at a premium. On the other hand, in the meetings with $k' > k$ and $\varepsilon' < \varepsilon$, or $k' < k$ and $\varepsilon' > \varepsilon$, the buyer bank searches faster than the seller bank, creating a speed discount to the bilateral Federal funds rate.

Search intensity. Given a quadratic value function, the equilibrium search intensity is the fixed point of

$$\Gamma_t(\varepsilon_t)(k) \equiv \arg\max_{\varepsilon\in[0,1]} \left\{ \int \left[\frac{k'-k}{2} V_t''(k)\right]^2 \underbrace{\frac{m(\varepsilon, \varepsilon_t(k'))}{\kappa[\varepsilon + \varepsilon_t(k')]}}_{\text{search efficiency}} \underbrace{\left[1 - \frac{V_t''(k) + V_t''(k')}{2\kappa[\varepsilon + \varepsilon_t(k')]}\right]^{-1}}_{\text{precaution-speed trade-off}} dF_t(k') \right\}, \tag{2.20}$$

Using Proposition 2.1, $\{\Omega(S_t, F_t), \succeq_s\}$ is a complete lattice since

$$X(k, k', \varepsilon, \varepsilon') = 1 - \frac{V_t''(k) + V_t''(k')}{2\kappa(\varepsilon + \varepsilon')} \geq 1,$$

and

$$\theta_\kappa(\varepsilon) = \theta_m(\varepsilon) = 1.$$

Thus, the equilibrium search profile can be ranked.

Proposition 2.4 ***(Multiplicity)*** *$\varepsilon_t(k) = 0\ \forall k$ is always an equilibrium search profile. Also,*

given V_t and F_t, the number of equilibrium search profiles at t, i.e., $|\Omega(S_t, F_t)|$, is greater than or equal to 1. When $|\Omega(S_t, F_t)| > 1$, $\varepsilon_t(k) = 1\ \forall k$ is the largest equilibrium search profiles.

We refer the smallest equilibrium search profile $\varepsilon_t(k) = 0$ as the frozen equilibrium. Similarly, we refer the largest equilibrium search profile, $\varepsilon_t(k) = 1$, if exists, as the most liquid equilibrium.

Verification. The following proposition verifies that the value function must be quadratic.

Proposition 2.5 ***(Closed form)*** *The value function of the largest equilibrium search profle admits a unique specification $V_t(k) = -H_t k^2 + E_t k + D_t$, where*

$$\dot{H}_t = rH_t - a_2 + \frac{1}{4}\frac{H_t^2}{\kappa\varepsilon_t + H_t}\left[(\lambda - \lambda_0)\varepsilon_t^2 + \lambda_0\right], \text{ where } H_T = A_2; \tag{2.21}$$

$$\dot{E}_t = rE_t - a_1 + \frac{K}{2}\frac{H_t^2}{\kappa\varepsilon_t + H_t}\left[(\lambda - \lambda_0)\varepsilon_t^2 + \lambda_0\right], \text{ where } E_T = A_1; \tag{2.22}$$

$$\dot{D}_t = rD_t - \frac{1}{4}\frac{H_t^2}{\kappa\varepsilon_t + H_t}\left[(\lambda - \lambda_0)\varepsilon_t^2 + \lambda_0\right]\int k'^2 dF_t(k'), \text{ where } D_T = 0. \tag{2.23}$$

The largest equilibrium search profile is $\varepsilon_t(k) = \varepsilon_t \in \{0, 1\}$. The Federal funds purchased $q_t(k, k', \varepsilon, \varepsilon')$ and the Federal funds rate $\rho_t(k, k', \varepsilon, \varepsilon')$ are given by

$$q_t(k, k', \varepsilon, \varepsilon') = \frac{H_t(k' - k)}{\kappa(\varepsilon + \varepsilon') + 2H_t}, \tag{2.24}$$

$$1 + \rho_t(k, k', \varepsilon, \varepsilon') = e^{r(T+\Delta-t)}\left[E_t - H_t(k + k') - \frac{\kappa(\varepsilon - \varepsilon')}{2}q_t(k, k', \varepsilon, \varepsilon')\right]. \tag{2.25}$$

2.5.1 The most liquid equilibrium

In the rest of this section, we focus on the largest equilibrium, which is referred as the most liquid equilibrium.

Proposition 2.6 *Define*

$$\eta \equiv \kappa \left[\frac{\lambda}{2(\lambda - \lambda_0)} - 1 \right].$$

The equilibrium search profile in the most active equilibrium is given by

$$\varepsilon_t(k) = \begin{cases} 1, & \text{if } V_t''(k) \leq -2\eta; \\ 0, & \text{otherwise.} \end{cases} \tag{2.26}$$

Given Proposition 2.6, the following lemma solves the path of equilibrium search profile in the most liquid equilibrium.

Lemma 2.2 *Define*

$$\begin{aligned} \mu_1 &\equiv \frac{1}{2r + \frac{\lambda}{2}} \left\{ -(\kappa r - a_2) - \left[(\kappa r - a_2)^2 + a_2 \kappa (4r + \lambda) \right]^{0.5} \right\}, \\ \mu_2 &\equiv \frac{1}{2r + \frac{\lambda}{2}} \left\{ -(\kappa r - a_2) + \left[(\kappa r - a_2)^2 + a_2 \kappa (4r + \lambda) \right]^{0.5} \right\}, \end{aligned}$$

$$\begin{aligned} \tau_1(H; A, u) &\equiv u - \frac{(\kappa + \mu_1) \log\left(\frac{A - \mu_1}{H - \mu_1}\right) - (\kappa + \mu_2) \log\left(\frac{A - \mu_2}{H - \mu_2}\right)}{\left(r + \frac{\lambda}{4}\right)(\mu_1 - \mu_2)}, \\ J(t; A, u) &\equiv \frac{a_2}{r + \frac{\lambda_0}{4}} + \left(A - \frac{a_2}{r + \frac{\lambda_0}{4}} \right) e^{-\left(r + \frac{\lambda_0}{4}\right)(u - t)}, \\ \tau_2(H; A, u) &\equiv u + \frac{1}{r + \frac{\lambda_0}{4}} \log\left(1 - \frac{H - A}{\frac{a_2}{r + \frac{\lambda_0}{4}} - A} \right). \end{aligned}$$

(a). Suppose $A_2 \geq \eta$.

(a-i). If $a_2 < \left(r - \frac{\lambda}{4} + \frac{\lambda_0}{2}\right)\eta$ *and* $\tau_1(\eta; A_2, T) > 0$, *then we have*

$$\varepsilon_t = \begin{cases} 1, & \text{if } t \geq \tau_1(\eta; A_2, T); \\ 0, & \text{otherwise.} \end{cases} \tag{2.27}$$

$$H_t = \begin{cases} \tau_1^{-1}(t; A_2, T), & \text{if } t \geq \tau_1(\eta; A_2, T); \\ J[t; \eta, \tau_1(\eta; A_2, T)], & \text{otherwise.} \end{cases} \tag{2.28}$$

(a-ii). Otherwise, we have $\varepsilon_t = 1$ *for all* $t \in [0, T]$ *and* $H_t = \tau_1^{-1}(t; A_2, T)$.

(b). Suppose $A_2 < \eta$.

(b-i). If $a_2 > \left(r + \frac{\lambda_0}{4}\right)\eta$ *and* $\tau_2(\eta; A_2, T) > 0$, *then we have*

$$\varepsilon_t = \begin{cases} 0, & \text{if } t > \tau_2(\eta; A_2, T); \\ 1, & \text{otherwise.} \end{cases} \tag{2.29}$$

$$H_t = \begin{cases} J(t; A_2, T), & \text{if } t \geq \tau_2(\eta; A_2, T); \\ \tau_1^{-1}(t; \eta, \tau_2(\eta; A_2, T)), & \text{otherwise.} \end{cases} \tag{2.30}$$

(b-ii). Otherwise, we have $\varepsilon_t = 0$ *for all* $t \in [0, T]$ *and* $H_t = J(t; A_2, T)$.

The above lemma shows that the path of equilibrium search intensity depends on the boundary value A_2, which is a function of the unconventional monetary policies $\left\{i^{ER}, i^{DW}, K\right\}$. The Federal funds market is not frozen ($\varepsilon_t = 1$) when A_2 is sufficiently large.

Having solved the time path of H_t, we are able to characterize the path of equilibrium reserve distribution as in the following lemma.

Lemma 2.3 *Given* H_t, *the reserve distribution under the largest equilibrium search profiles solves the following PDE:*

$$\dot{F}_t(k) = m(\varepsilon_t, \varepsilon_t)\left[\int F_t\left[2\left(1 + \frac{\kappa\varepsilon_t}{H_t}\right)k - \left(1 + \frac{2\kappa\varepsilon_t}{H_t}\right)k'\right]dF_t(k') - F_t(k)\right], \tag{2.31}$$

given the initial condition $F_0(k)$. *Denote the* n*-th moment of the reserve distribution at time*

t as $M_{n,t} \equiv \int k^n dF_t(k)$. *The moment function is given by the following ODE:*

$$\dot{M}_{n,t} = m(\varepsilon_t, \varepsilon_t) \left[\sum_{i=0}^{n} C_n^i \frac{(H_t)^{n-i} (H_t + 2\kappa\varepsilon_t)^i}{2^n (H_t + \kappa\varepsilon_t)^n} M_{n-i,t} M_{i,t} - M_{n,t} \right], \tag{2.32}$$

with $M_{0,t} = 1$, $M_{1,t} = K$ *and*

$$M_{2,t} = K^2 + \left(M_{2,0} - K^2\right) \exp\left[- \int_0^t m(\varepsilon_z, \varepsilon_z) \frac{H_z (H_z + 2\kappa\varepsilon_z)}{2 (H_z + \kappa\varepsilon_z)^2} dz \right]. \tag{2.33}$$

Thanks to Fourier transform, the model allows for an analytical expression for the paths of moments of the reserve distribution. In particular, equation (2.33) implies that the variance of the reserve distribution converges to zero at the speed of $m(\varepsilon_t, \varepsilon_t) \frac{H_t(H_t+2\kappa\varepsilon_t)}{2(H_t+\kappa\varepsilon_t)^2}$, which is endogenously determined. In particular, a higher H_t implies a faster speed of convergence.

2.5.2 Positive implications on liquidity

The closed-form solution allows us to obtain a set of measures on liquidity in analytical form. We list these measures in this section for possible quantitative analysis. The derivations of all the measures are provided in the Appendix .

Price impact. The price impact of a trade measures how much the Federal fund rate changes in response to a given Federal fund purchased. The higher the price impact, the more expensive to borrow reserve balances, reflecting lower liquidity. In the Walrasian benchmark, the price impact is always zero. Substituting the equilibrium search intensity in our model, the Federal fund rate can be log-linearized as

$$\rho_t(k, q) \cong \underbrace{r(T + \Delta - t)}_{\text{time effect}} + \underbrace{\log V_t'(k)}_{\text{bank fixed effect}} - \underbrace{\frac{\theta_{V,t}(k)}{1 - \omega_t}}_{\text{price impact}} \frac{q}{k}, \tag{2.34}$$

where $\theta_{V,t}(k)$ is the elasticity of value function and ω_t is the equilibrium precaution-speed trade-off:

$$\begin{aligned}\theta_{V,t}(k) &\equiv -\frac{V_t''(k)\,k}{V_t'(k)},\\ \omega_t &\equiv \left(1-\frac{\bar{V}_t''}{\kappa\varepsilon_t}\right)^{-1}.\end{aligned}$$

The price impact depends on the ratio between the elasticity of value function and the precaution-speed trade-off.

Return reversal. If the Federal funds market is liquid, the price impact is transistory and the Federal fund rate will swiftly reverse to the mean. The return reversal measures how swift the Federal fund rate stablizes disturbances. In the Walrasian benchmark, the return reversal is always infinity. In our model, the dynamics of the Federal fund rate is given by

$$\frac{d}{dt}\left[\rho_t(k,k')-\varrho_t\right]=-\underbrace{\left[\frac{a_2}{H_t}-\frac{1}{4}\frac{H_t}{\kappa\varepsilon_t+H_t}\left[(\lambda-\lambda_0)\,\varepsilon_t^2+\lambda_0\right]\right]}_{\text{return reversal}}\left[\rho_t(k,q)-\varrho_t\right],$$

where ϱ_t is the average Federal funds rate defined by $\varrho_t \equiv \int\int \rho_t(k,k')\,dF_t(k')\,dF_t(k)$. Note that the value of $V_t''(k)$ and the search intensity both control the speed of return reversal.

Price dispersion. The law of one price tends to apply when the Federal fund market is extremely liquid. The price dispersion measures the prevalence of arbitrage opportunity arise of the search friction. In the Walrasian benchmark, the price dispersion is always zero. In our model, the price dispersion is given by

$$\underbrace{\frac{\sigma_{\rho,t}}{\sigma_{k,t}}}_{\text{price dispersion}}=\sqrt{2}e^{r(T+\Delta-t)}H_t,$$

where $\sigma_{\rho,t}$ is the standard deviation of Federal fund rate and $\sigma_{k,t}$ is the standard deviation of reserve balances. Since the Federal fund rates are more dispersed when banks hold more dispersed reserve balances, we normalize the price dispersion with the standard deviation of reserve balances.

Intermediation markup. Recall that banks intermediate by purchasing Federal funds to sell. Intermediation is not risk-free as the bank exposes itself to the risk of selling Federal funds at a lower price than the purchasing price. The rate spread is the between the expected Federal fund rate of the selling leg and the realized Federal fund rate of the purchasing leg:

$$\Delta_{\rho,t}(k,q) \equiv \int \rho_t(k+q,k')\, dF_t(k') - \rho_t(k,q).$$

The intermediation markup measures the change in the rate spread in response to the size of the intermediation trade. In our model, the intermediation markup is given by

$$\underbrace{\frac{\partial \Delta_{\rho,t}(k,q)}{\partial q}}_{\text{intermediation markup}} = e^{r(T+\Delta-t)}\left(2\kappa\varepsilon_t + H_t\right).$$

Utilization rate of trade opportunities. The total trade opportunities in this economy is

$$TO_t = \int_k \int_{k'\geq k} \frac{k'-k}{2} dF_t(k')\, dF_t(k).$$

The utilitzation rate of trade opportunities measure how fast the trade opportunities are realized. In Afonso and Lagos (2012), the utilization rate is the exogeneous matching rate. In our model, the utilization rate is

$$UR_t = \frac{\int_k \int_{k'\geq k} m(\varepsilon_t,\varepsilon_t)\, q_t(k,k',\varepsilon_t,\varepsilon_t)\, dF_t(k')\, dF_t(k)}{TO_t} = \frac{H_t\left[(\lambda-\lambda_0)\,\varepsilon_t^2+\lambda_0\right]}{\kappa\varepsilon_t + H_t}.$$

Peak of trades. According to Proposition 2.5, the search decision is summarized by whether or not the condition $H_t \geq \eta$ is satisfied.

Extensive margins. The measure of intermediating banks and the amount of intermediated reserves are characterized by ODEs. Denote

$$\begin{aligned}
P_t^b(k) &\equiv \Pr\{q_z(k_z, k_z', \varepsilon_z, \varepsilon_z) > 0 | k_t = k, z \geq t\}, \\
P_t^s(k) &\equiv \Pr\{q_z(k_z, k_z', \varepsilon_z, \varepsilon_z) < 0 | k_t = k, z \geq t\}, \\
P_t^{tr}(k) &\equiv \Pr\{q_z(k_z, k_z', \varepsilon_z, \varepsilon_z) \neq 0 | k_t = k, z \geq t\}, \\
P_t^{\text{int}}(k) &\equiv \Pr\{q_z(k_z, k_z', \varepsilon_z, \varepsilon_z) > 0, q_{z'}(k_{z'}, k_{z'}', \varepsilon_{z'}, \varepsilon_{z'}) < 0 | k_t = k, z \geq t, z' \geq t\},
\end{aligned}$$

where $P_t^b(k)$ is the probability that a k-bank will borrow reserves during the remaining time $[t, T]$, and similiarly $P_t^s(k)$ is the corresponding probability of lending reserves, $P_t^{tr}(k)$ the corresponding probability of trading reserves, and $P_t^{\text{int}}(k)$ the corresponding probability of intermediating reserves. By the law of large number, $P^b \equiv \int P_0^b(k)\, dF(k)$ is the measure of banks that borrow in the Federal funds market. Similarly, $P^s \equiv \int P_0^s(k)\, dF(k)$ is the measure of lending banks, $P^{tr} \equiv \int P_0^{tr}(k)\, dF(k)$ is the measure of trading banks, and $P^{\text{int}} \equiv \int P_0^{\text{int}}(k)\, dF(k)$ is the measure of intermediating banks. By definition we have $P^{\text{int}} = P^b + P^s - P^{tr}$.

The laws of motion for the measures of trading banks, lending banks, borrowing banks,

and intermediating banks are given by

$$\begin{aligned}
0 &= \dot{P}_t^{tr}(k) + m_t\left[1 - P_t^{tr}(k)\right], \\
0 &= \dot{P}_t^{b}(k) + m_t\left[1 - F_t(k)\right]\left[1 - P_t^{b}(k)\right] + m_t\int_{k'\leq k}\left[P_t^{b}\left[k + q_t(k,k')\right] - P_t^{b}(k)\right]dF_t(k'), \\
0 &= \dot{P}_t^{s}(k) + m_t F_t(k)\left[1 - P_t^{s}(k)\right] + m_t\int_{k'\geq k}\left[P_t^{s}\left[k + q_t(k,k')\right] - P_t^{s}(k)\right]dF_t(k'), \\
0 &= \dot{P}_t^{\text{int}}(k) + m_t\int_{k'\leq k}\left[P_t^{b}\left[k + q_t(k,k')\right] - P_t^{\text{int}}(k)\right]dF_t(k') \\
&\quad + m_t\int_{k'\geq k}\left[P_t^{s}\left[k + q_t(k,k')\right] - P_t^{\text{int}}(k)\right]dF_t(k'),
\end{aligned}$$

where the boundary condition is given by $P_T^{tr}(k) = P_T^{b}(k) = P_T^{s}(k) = 0$. Note that only $P_t^{tr}(k)$ has a closed-form solution:

$$P_t^{tr}(k) = 1 - \exp\left[-\int_t^T m(\varepsilon_z, \varepsilon_z)\,dz\right]. \tag{2.35}$$

Intensive margins. We define two measures of intensive margins for trade. The first measure is the cumulated amount of absolute trade volume from time t to T for a bank with k units of reserve balances at time t:

$$Q_t(k) \equiv \mathbb{E}\sum_{t_i\in[t,T]}\left|q_{t_i}(k_{t_i}, k')\right| \text{ s.t. } k_t = k. \tag{2.36}$$

The individual absolute trades follows

$$\dot{Q}_t(k) = -m(\varepsilon_t, \varepsilon_t)\left[\int_{k'}\left|q_t(k,k')\right|dF_t(k') + \int_{k'}Q_t\left(k + q_t(k,k')\right)dF_t(k') - Q_t(k)\right]. \tag{2.37}$$

By summing up $Q_t(k)$ we can obtain the aggregate volume of absolute trades

$$Q_t \equiv \int Q_t(k)\,dF_t(k).$$

The aggregate absolute trades follows the following ODE:

$$\dot{Q}_t = -m\left(\varepsilon_t, \varepsilon_t\right) \frac{H_t}{\kappa\varepsilon_t + H_t} \int\int \frac{|k' - k|}{2} dF_t\left(k'\right) dF_t\left(k\right).$$

Thus the total trade volume is

$$Q = \int_0^T \frac{m\left(\varepsilon_t, \varepsilon_t\right) H_t}{2\left(\kappa\varepsilon_t + H_t\right)} \left(\int\int |k' - k| \, dF_t\left(k'\right) dF_t\left(k\right) \right) dt.$$

The second measure is the net trade volume, i.e. the net Federal funds purchased. We define the expected amount of net trades from time t to T of a bank holding k units of reserve balances at time t as

$$L_t\left(k\right) \equiv \mathbb{E} \sum_{t_i \in [t,T]} q_{t_i}\left(k_{t_i}, k'\right) \text{ s.t. } k_t = k.$$

The aggregate absolute net trade is defined as

$$L \equiv \int \left|L_0\left(k\right)\right| dF_0\left(k\right).$$

Note that the individual net trade admits a closed-form solution:

$$L_t\left(k\right) = \left\{ 1 - \exp\left[- \int_t^T \frac{m\left(\varepsilon_z, \varepsilon_z\right) H_z}{2\left(\kappa\varepsilon_z + H_z\right)} dz \right] \right\} \left(K - k\right). \tag{2.38}$$

We can think of $L_t\left(k\right)$ as the net trade volume of bank k who contacts bank K at intensity $m\left(\varepsilon_t, \varepsilon_t\right)$. Thanks to the closed-form solution, we also derive the comparative statics of $L_t\left(k\right)$ on policy parameters in Section 2.5.3. Given the individual net trade volume, the aggregate volume of the absolute net trade is

$$L \equiv \int \left|L_0\left(k\right)\right| dF_0\left(k\right) = \left\{ 1 - \exp\left[- \int_0^T \frac{m\left(\varepsilon_t, \varepsilon_t\right) H_t}{2\left(\kappa\varepsilon_t + H_t\right)} dt \right] \right\} \int \left|K - k\right| dF_0\left(k\right).$$

Given the aggregate volume of absolute trade and net trade, we define the level of intermediation and fraction of intermediation as

$$\begin{aligned} \text{Int} &= Q - L, \\ \text{IntR} &= \frac{Q - L}{Q}. \end{aligned}$$

Federal fund rate. The average Federal fund rate at τ is given by

$$\begin{aligned} 1 + \varrho_t &= \int\int \left[1 + \rho_t\left(k, k'\right)\right] dF_t\left(k'\right) dF_t\left(k\right) = e^{r(T+\Delta-t)}\left[E_t - 2H_t K\right] \\ &= e^{r\Delta}\left[1 + \gamma + i^{ER} + \frac{k_+ - 1}{k_+ - k_-}\Delta i\right] - \frac{2a_2 K - a_1}{r}\left[e^{r(T+\Delta-t)} - e^{r\Delta}\right], \end{aligned}$$

where $\Delta i = i^{DW} - i^{ER}$ is the policy rate spread. The range of the Federal funds rates is given by $1 + \rho_t\left(k, k'\right) \in \left[1 + \rho_t^{\min}, 1 + \rho_t^{\max}\right]$, where

$$\begin{aligned} 1 + \rho_t^{\min} &= e^{r(T+\Delta-t)}\left[E_t - 2H_t k_{\max}\right], \\ 1 + \rho_t^{\max} &= e^{r(T+\Delta-t)}\left[E_t - 2H_t k_{\min}\right]. \end{aligned}$$

2.5.3 Comparative statics

This section provides the comparative statics of the closed-form solutions to policy and technology parameters. We focus on the comparative statics where T is sufficiently small, corresponding to the fact that the Federal funds market is usually active during the last 2.5 hrs of a trading session. Based on the characterization of equilibrium paths in Lemma 2.2, we discuss the comparative statics of two cases. The first case is that banks search at the beginning of the trading session, and the second case is that banks search when the time gets close to the end of trading session. The following Proposition summarizes the comparative statics for the first case.

Proposition 2.7 *Suppose* $A_2 < \eta$, $a_2 > \left(r + \frac{\lambda_0}{4}\right)\eta$, $\tau_2(\eta; A_2, T) > 0$, *and* T *is sufficiently small. The comparative statics of the length of search,* $\tau_2(\eta; A_2, T)$, *the amount of Federal funds purchased,* $q_t(k, k')$, *net Federal funds purchase,* $L_0(k)$ *and its derivative* $L_0'(k)$, *and the bilateral Federal fund rates,* $\rho_t(k, k')$, *with respect to* i^{ER}, i^{DW}, κ, λ_0, λ *and* K, *are given by the following table*

	τ_2	$\lvert q_t\rvert$	$L_0(k)$	$L_0'(k)$	$\rho_t(k, k')$
i^{ER}	$-$	$-$	$sgn(K-k)$	$-$	$+(-)$ *for* $k + k' > (<)\, \hat{K}_t(k_-)$
i^{DW}	$+$	$+$	$sgn(k-K)$	$+$	$+(-)$ *for* $k + k' < (>)\, \hat{K}_t(k_+)$
K	$-$	$-$	$+(-)$ *for small (large)* k	$-$	$+(-)$ *for* $k + k' > (<)\, \hat{K}_t(\zeta_t)$
κ	$-$	$-$	$sgn(K-k)$	$-$	$+(-)$ *for* $k + k' < (>)\, 2K$
λ_0	$-$	$-$	$sgn(K-k)$	$-$	$+(-)$ *for* $k + k' > (<)\, 2K$
λ	$+$	$+$	$sgn(k-K)$	$+$	$+(-)$ *for* $k + k' < (>)\, 2K$

where

$$\zeta_t = \int_t^T e^{r(T-s)} \frac{\left[\left(\lambda - \lambda_0\right)\varepsilon_s^2 + \lambda_0\right] H_s^2}{4A_2\left(\kappa\varepsilon_s + H_s\right)} ds,$$

$$\hat{K}_t(k^w) = 2K \times \frac{k^w - 1 + \exp\left[-\frac{\lambda_0}{4}\left(T - t - \left(\tau_2(\eta; A_2, T) - t\right)^+\right)\right] - M\left(\left(\tau_2(\eta; A_2, T) - t\right)^+\right)\exp(rT)}{\exp\left[-\frac{\lambda_0}{4}\left(T - t - \left(\tau_2(\eta; A_2, T) - t\right)^+\right)\right] - M\left(\left(\tau_2(\eta; A_2, T) - t\right)^+\right)\exp(rT)},$$

and

$$M(u) = \frac{\partial \tau_2(\eta; A_2, T)}{\partial A_2} \int_{\tau_2(\eta; A_2, T) - u}^{\tau_2(\eta; A_2, T)} e^{-rs} \left\{\left(r + \frac{\lambda}{4}\right)\left[1 - \frac{(\kappa + \mu_1)(\kappa + \mu_2)}{(\kappa + H_s)^2}\right]\right\}\left(-\dot{H}_s\right) ds,$$

and $(x)^+ \equiv \max\{x, 0\}$.

The following proposition summarizes the comparative statics for the second case.

Proposition 2.8 *Suppose* $A_2 \geq \eta$, $a_2 < \left(r - \frac{\lambda}{4} + \frac{\lambda_0}{2}\right)\eta$, $\tau_1(\eta; A_2, T) > 0$ *and* λ, λ_0/λ *and* T *are sufficiently small. The comparative statics of the length of search,* $T - \tau_1(\eta; A_2, T)$,

the amount of Federal funds purchased, $q_t(k,k')$, net Federal funds purchase, $L_0(k)$ and its derivative $L_0'(k)$, and the Federal fund rates, $\rho_t(k,k')$, with respect to i^{ER}, i^{DW}, κ, λ_0, λ and K, are given by the following table

	$T-\tau_1$	$\lvert q_t\rvert$	$L_0(k)$	$L_0'(k)$	$\rho_t(k,k')$
i^{ER}	$-$	$-$	$sgn(k-K)$	$+$	$+(-)$ *for* $k+k'>(<)\,\tilde{K}_t(k_-)$
i^{DW}	$+$	$+$	$sgn(K-k)$	$-$	$+(-)$ *for* $k+k'<(>)\,\tilde{K}_t(k_+)$
K	$-$	$-$	$+(-)$ *for large (small)* k	$+$	$+(-)$ *for* $k+k'>(<)\,\tilde{K}_t(\zeta_t)$
κ	$-$	$-$	$sgn(k-K)$	$+$	$+(-)$ *for* $k+k'<(>)\,2K$
λ_0	$-$	0	$sgn(K-k)$	$-$	$+(-)$ *for* $k+k'>(<)\,2K$
λ	$+$	$-$	$sgn(K-k)$	$-$	$+(-)$ *for* $k+k'>(<)\,2K$

where

$$\zeta_t=\int_t^T e^{r(T-s)}\frac{\left[(\lambda-\lambda_0)\,\varepsilon_s^2+\lambda_0\right]H_s^2}{4A_2\left(\kappa\varepsilon_s+H_s\right)}ds,$$

$$\tilde{K}_t(k^w)=2K\cdot\frac{k^w-\tilde{M}\left(T-\tau_1-(t-\tau_1)^+\right)\exp(rT)-1\{t<\tau_1\}\,e^{r(T-\tau_1)}\frac{(\lambda-\lambda_0)\eta}{4}\frac{\partial\tau_1}{\partial A_2}}{1-\tilde{M}\left(T-\tau_1-(t-\tau_1)^+\right)\exp(rT)-1\{t<\tau_1\}\,e^{r(T-\tau_1)}\frac{(\lambda-\lambda_0)\eta}{4}\frac{\partial\tau_1}{\partial A_2}},$$

and

$$\tilde{M}(u)=\frac{\lambda}{4}\frac{\partial\tau_1(\eta;A_2,T)}{\partial A_2}\int_{T-u}^T e^{-rs}\frac{H_s(2\kappa+H_s)}{(\kappa+H_s)^2}\left(-\dot{H}_s\right)ds,$$

and $(x)^+\equiv\max\{x,0\}$.

The above propositions show that the comparative statics may differ in different cases and depend on parameters. In particular, the comparative statics on the length of search are consistent for the two cases. The length of search decreases in IOER and aggregate excess reserves, implying the disintermediation effect on the extensive margin. However, there is a trade-off at the intensive margin. The bilateral trade size when the search intensity is one is always smaller than the trade size when the search intensity is zero. Thus a shorter length of search does not necessarily imply a lower volume of transaction. The disintermediation effect on the intensive margin occurs only if the reduction in trade size when the search intensity

is one dominates. This is the case in Proposition 2.8. Note that in this case, banks' search intensity is one when the time is close to the end of the market. This is consistent with the empirical observation that the daily Federal funds market is usually active during 4pm to 6:30pm. Moreover, Proposition 2.8 also produces the comparative statics of net Federal funds purchase that are consistent with the empirical evidence.

2.5.4 Constrained efficiency

In this section we discuss the constrained efficiency of the closed-form model. The following proposition characterizes the necessary conditions for the planner's problem.

Proposition 2.9 *A solution to the planner's problem is a path for the distribution balances,* $F_t^p(k)$*, a path for the continuum of co-states associated with the law of motion for the distribution of balances,* $\mathbf{V}_t^p = \{V_t^p(k)\}_{k\in\mathbb{K}}$*, a path for the individual search intensity profile* $\{\varepsilon_t^p(k)\}_{k\in\mathbb{K}}$*, and a path for the bilateral reallocation volume,* $\{q_t^p(k,k')\}_{k,k'\in\mathbb{K}}$*. The necessary conditions for optimality are*

$$\begin{aligned} & rV_t^p(k) \qquad (2.39)\\ = \;& \dot{V}_t^p(k) + u(k) \\ & + \max_{\varepsilon\in[0,1]} \int_{k'} \max_{\substack{q\in\mathbb{R} \\ k+q,k'-q\in\mathbb{K}}} \left\{ \begin{array}{c} V_t^p(k+q) - V_t^p(k) - \chi(\varepsilon,q) \\ +V_t^p(k'-q) - V_t^p(k) - \chi(\varepsilon_t^p(k'),-q) \end{array} \right\} m(\varepsilon,\varepsilon_t^p(k'))\, dF_t^p(k') \end{aligned}$$

for all $(k,t)\in\mathbb{K}\times[0,T]$*, with*

$$V_T^p(k) = U(k) \;\; \textit{for all } k\in\mathbb{K}, \qquad (2.40)$$

with the path for $F_t^p(k)$ *given by*

$$\dot{F}_t^p(k^w) = \left\{ \begin{array}{c} \int_{k>k^w} \int m\left[\varepsilon_t^p(k), \varepsilon_t^p(k')\right] 1\left\{k + q_t^p(k,k') \leq k^w\right\} dF_t^p(k') dF_t^p(k) \\ -\int_{k\leq k^w} \int m\left[\varepsilon_t^p(k), \varepsilon_t^p(k')\right] 1\left\{k + q_t^p(k,k') > k^w\right\} dF_t^p(k') dF_t^p(k) \end{array} \right\}, \quad (2.41)$$

where $F_0^p(k^w) = F_0(k^w)$.

Note that the maximization problem in the planner's HJB (2.39) is different from the counterpart in the equilibrium, creating the inefficiency of the equilibrium allocation. The difference is due to a composition externality typical of *ex post* bargaining environments, as discussed by Afonso and Lagos (2015b). An individual bank internalizes only half the surpluses that her trades create. As a result, she does not internalize fully the social benefit as well as social cost that arise from the fact that having her in the current reserve holding k increases the meeting intensity of all other banks with a bank of reserve k. Different from Afonso and Lagos (2015b), the post trading reserve holdings of a bank k and a bank k' is a weighted average of k and k' due to the endogenous transaction costs, and the weight is dependent on the composition externality.

Following the method we use in the equilibrium analysis, we guess and verify that

$$V_t^p(k) = -H_t^p k^2 + E_t^p k + D_t^p. \quad (2.42)$$

If the optimal reallocation rule $q_t^p(k,k')$ is non-zero, it satisfies

$$q_t^p(k,k') = \frac{H_t^p(k'-k)}{2H_t^p + \kappa\left[\varepsilon_t^p(k) + \varepsilon_t^p(k')\right]}. \quad (2.43)$$

The bilateral surplus is

$$S_t^p(k,k') = \frac{\left[H_t^p(k'-k)\right]^2}{2H_t^p + \kappa\left[\varepsilon_t^p(k) + \varepsilon_t^p(k')\right]}. \quad (2.44)$$

The optimal search intensity satisfies

$$\Gamma_t^p\left(\varepsilon_t^p\right)(k) \equiv \arg\max_{\varepsilon\in[0,1]} \int_{k'} S_t^p\left(k,k',\varepsilon,\varepsilon_t^p\left(k'\right)\right) m\left(\varepsilon,\varepsilon_t^p\left(k'\right)\right) dF_t^p\left(k'\right) \tag{2.45}$$

Therefore the HJB (2.39) simplifies to

$$rV_t^p(k) = \dot{V}_t^p(k) + u(k) + \frac{\left(H_t^p\right)^2}{2}\frac{(\lambda-\lambda_0)\left(\bar{\varepsilon}_t^p\right)^2+\lambda_0}{\kappa\bar{\varepsilon}_t^p + H_t^p}\int_{k'}\left(k-k'\right)^2 dF_t^p\left(k'\right). \tag{2.46}$$

Matching coefficients yields

$$\dot{H}_t^p = rH_t^p - a_2 + \frac{\left(H_t^p\right)^2}{2}\frac{(\lambda-\lambda_0)\left(\bar{\varepsilon}_t^p\right)^2+\lambda_0}{\kappa\bar{\varepsilon}_t^p + H_t^p}, \text{ with } H_T^p = A_2. \tag{2.47}$$

$$\dot{E}_t^p = rE_t^p - a_1 + K\left(H_t^p\right)^2\frac{(\lambda-\lambda_0)\left(\bar{\varepsilon}_t^p\right)^2+\lambda_0}{\kappa\bar{\varepsilon}_t^p + H_t^p}, \text{ with } E_T^p = A_1. \tag{2.48}$$

$$\dot{D}_t^p = rD_t^p - \frac{\left(H_t^p\right)^2}{2}\frac{(\lambda-\lambda_0)\left(\bar{\varepsilon}_t^p\right)^2+\lambda_0}{\kappa\bar{\varepsilon}_t^p + H_t^p}\int_{k'}\left(k'\right)^2 dF_t^p\left(k'\right), \text{ with } D_T^p = 0. \tag{2.49}$$

Note that the initial-value problem (2.47) of H_t^p has the same functional form as that of H_t, except that the parameters in the former are all doubled compared to the latter. Therefore, Lemma 1 and 2 also apply to H_t^p, and we can get the property of time path of reallocation that is similar to Proposition 8. In particular, we define

$$\eta^p = \kappa\left[\frac{\lambda}{2(\lambda-\lambda_0)} - 1\right] = \eta.$$

This implies that the switching point of the time path of reallocation has the same cutoff value of H, but the cutoff time τ can be different since the "search intensity" is higher.

Lemma 2.4 *Define*

$$\begin{aligned}
\mu_1^p &\equiv \frac{1}{2r+\lambda}\left\{-(\kappa r - a_2) - \left[(\kappa r - a_2)^2 + a_2\kappa(4r+2\lambda)\right]^{0.5}\right\}, \\
\mu_2^p &\equiv \frac{1}{2r+\lambda}\left\{-(\kappa r - a_2) + \left[(\kappa r - a_2)^2 + a_2\kappa(4r+2\lambda)\right]^{0.5}\right\},
\end{aligned}$$

$$\begin{aligned}
\tau_1^p(H;A,u) &\equiv u - \frac{(\kappa+\mu_1^p)\log\left(\frac{A-\mu_1^p}{H-\mu_1^p}\right) - (\kappa+\mu_2^p)\log\left(\frac{A-\mu_2^p}{H-\mu_2^p}\right)}{\left(r+\frac{\lambda}{2}\right)(\mu_1^p-\mu_2^p)}, \\
J^p(t;A,u) &\equiv \frac{a_2}{r+\frac{\lambda_0}{2}} + \left(A - \frac{a_2}{r+\frac{\lambda_0}{2}}\right)e^{-\left(r+\frac{\lambda_0}{2}\right)(u-t)}, \\
\tau_2^p(H;A,u) &\equiv u + \frac{1}{r+\frac{\lambda_0}{2}}\log\left(1 - \frac{H-A}{\frac{a_2}{r+\frac{\lambda_0}{2}} - A}\right).
\end{aligned}$$

(a) Suppose $A_2 \geq \eta^p$.

(a-i). If $a_2 < \left(r - \frac{\lambda}{2} + \lambda_0\right)\eta$ *and* $\tau_1^p(\eta^p;A_2,T) > 0$, *then we have*

$$\varepsilon_t^p = \begin{cases} 1, & \text{if } t \geq \tau_1^p(\eta;A_2,T); \\ 0, & \text{otherwise.} \end{cases} \tag{2.50}$$

$$H_t^p = \begin{cases} \left(\tau_1^p\right)^{-1}(t;A_2,T), & \text{if } t \geq \tau_1^p(\eta;A_2,T); \\ J^p\left[t;\eta,\tau_1^p(\eta;A_2,T)\right], & \text{otherwise.} \end{cases} \tag{2.51}$$

(a-ii). Otherwise, we have $\varepsilon_t^p = 1$ *for all* $t \in [0,T]$ *and* $H_t = \left(\tau_1^p\right)^{-1}(t;A_2,T)$.

(b). Suppose $A_2 < \eta^p$.

(b-i). If $a_2 > \left(r + \frac{\lambda_0}{2}\right)\eta$ *and* $\tau_2^p(\eta^p;A,T) > 0$, *then we have*

$$\varepsilon_t^p = \begin{cases} 0, & \text{if } t > \tau_2^p(\eta;A,T); \\ 1, & \text{otherwise.} \end{cases} \tag{2.52}$$

$$H_t^p = \begin{cases} J^p\left(t; A_2, T\right), \ \textit{if } t \geq \tau_2^p\left(\eta; A, T\right); \\ \left(\tau_1^p\right)^{-1}\left(t; \eta, \tau_2^p\left(\eta; A, T\right)\right), \ \textit{otherwise.} \end{cases} \tag{2.53}$$

(b-ii). Otherwise, we have $\varepsilon_t^p = 0$ *for all* $t \in [0, T]$ *and* $H_t = J^p\left(t; A_2, T\right)$.

This proposition implies that during the trading session, part of inefficiency can come from extensive margin, i.e. the timing and time length of reallocation, and the rest can com from intensive margin, i.e. the size of reserve reallocation. The following proposition characterizes this result for both cases in the above proposition.

Proposition 2.10 *For case (a-i) and (b-i) in Lemma 2.4, there are both inefficiencies on extensive and intensive margin. The active reallocation time length is shorter than equilibrium solution, and the reallocation size is smaller in the constrained efficiency solution.*

For case (a-ii) and (b-ii) in Lemma 2.4, there is no efficiency loss on extensive margin, but the reallocation size in a meeting is smaller in the constrained efficinecy solution.

Although the matching function implies complementarity between banks' search, banks are not supposed to under-search due to the positive externality. Instead, banks are actually trading too much, in terms of extensive and intensive margins, in the equilibrium than the constrained optimuml. Here is the reason. In the Federal funds market banks rely on bilateral trades to achieve their target levels of reserve holding. But trades in the OTC market is opportunistic, thanks to the search frictions, so banks tend to over-trade whenever they have a chance. Similarly, banks tend to search longer to compensate the search frictions. In sum, banks are trading too much in the equilibrium because of the precautionary motive, amplified by the search friction.

In the equilibrium, banks want to trade to the middle of the distribution - it is clear in the $k \in \{0, 1, 2\}$ model. In the constrained optimum, being the "middle bank" is not that good to the economy. The contribution from both ends of the distribution is much higher than the middle, as they create more trade surplus to their counterparties, which is not

internalized. To the individual bank and the planner, the motivation of trade is to narrow the dispersion of reserves, but the dispersion is more costly to the individual bank than to the planner. Therefore, banks have more incentive to trade to the middle than the planner. It results in over-search and over-intermediation in the equilibrium.

Our results are novel in the literature. Farboodi et al. (2017) obtains the similar argument, but the matching function in their model exhibits negative congestion externality, so agents oversearch in a steady state equilibrium. Our model has no congestion externality: matching function is increasing returns to scale, and the trading game is supermodular.

We can also obtain similar comparative statics of the constrained efficiency allocation as the equilibrium solution. The following proposition summarizes the results.

Proposition 2.11 *(1) Suppose* $A_2 < \eta^p$, $a_2 > \left(r + \frac{\lambda_0}{4}\right)\eta^p$, $\tau_2^p\left(\eta^p; A_2, T\right) > 0$, *and* T *is sufficiently small. The comparative statics of the length of search,* $\tau_2^p\left(\eta^p; A_2, T\right)$, *the amount of Federal funds purchased,* $q_t^p\left(k, k'\right)$, *net Federal funds purchase,* $L_0^p\left(k\right)$ *and its derivative* $L_0^{p\prime}\left(k\right)$, *and the bilateral Federal fund rates,* $\rho_t^p\left(k, k'\right)$, *with respect to* i^{ER}, i^{DW}, κ, λ_0, λ *and* K, *are given by the following table*

	τ_2^p	$\lvert q_t^p\rvert$	$L_0^p\left(k\right)$	$L_0^{p\prime}\left(k\right)$	$\rho_t^p\left(k,k'\right)$
i^{ER}	$-$	$-$	$sgn(K-k)$	$-$	$+\left(-\right)$ *for* $k+k' > \left(<\right) \hat{K}_t^p\left(k_-\right)$
i^{DW}	$+$	$+$	$sgn(k-K)$	$+$	$+\left(-\right)$ *for* $k+k' < \left(>\right) \hat{K}_t^p\left(k_+\right)$
K	$-$	$-$	$+\left(-\right)$ *for small (large)* k	$-$	$+\left(-\right)$ *for* $k+k' > \left(<\right) \hat{K}_t^p\left(\zeta_t\right)$
κ	$-$	$-$	$sgn(K-k)$	$-$	$+\left(-\right)$ *for* $k+k' < \left(>\right) 2K$
λ_0	$-$	$-$	$sgn(K-k)$	$-$	$+\left(-\right)$ *for* $k+k' > \left(<\right) 2K$
λ	$+$	$+$	$sgn(k-K)$	$+$	$+\left(-\right)$ *for* $k+k' < \left(>\right) 2K$

where

$$\zeta_t^p = \int_t^T e^{r(T-s)} \frac{\left[\left(\lambda-\lambda_0\right)\varepsilon_s^2 + \lambda_0\right]\left(H_s^p\right)^2}{2A_2\left(\kappa\varepsilon_s^p + H_s^p\right)} ds,$$

$$\hat{K}_t^p\left(k^w\right)=2K\times$$
$$\frac{k^w-1+\exp\left[-\frac{\lambda_0}{2}\left(T-t-\left(\tau_2\left(\eta;A_2,T\right)-t\right)^+\right)\right]-M^p\left(\left(\tau_2^p\left(\eta^p;A_2,T\right)-t\right)^+\right)\exp\left(rT\right)}{\exp\left[-\frac{\lambda_0}{2}\left(T-t-\left(\tau_2\left(\eta;A_2,T\right)-t\right)^+\right)\right]-M^p\left(\left(\tau_2^p\left(\eta^p;A_2,T\right)-t\right)^+\right)\exp\left(rT\right)},$$

and

$$M^p\left(u\right)=\frac{\partial\tau_2^p\left(\eta^p;A_2,T\right)}{\partial A_2}\int_{\tau_2^p\left(\eta^p;A_2,T\right)-u}^{\tau_2^p\left(\eta^p;A_2,T\right)}e^{-rs}\left\{\left(r+\frac{\lambda}{4}\right)\left[1-\frac{\left(\kappa+\mu_1^p\right)\left(\kappa+\mu_2^p\right)}{\left(\kappa+H_s^p\right)^2}\right]\right\}\left(-\dot{H}_s^p\right)ds,$$

and $\left(x\right)^+\equiv\max\left\{x,0\right\}$.

(2) Suppose $A_2\geq\eta^p$, $a_2<\left(r-\frac{\lambda}{4}+\frac{\lambda_0}{2}\right)\eta^p$, $\tau_1^p\left(\eta^p;A_2,T\right)>0$ *and* λ, λ_0/λ *and* T *are sufficiently small. The comparative statics are given by the following table*

	$T-\tau_1^p$	$\left\|q_t^p\right\|$	$L_0^p\left(k\right)$	$L_0^{p\prime}\left(k\right)$	$\rho_t^p\left(k,k'\right)$
i^{ER}	$-$	$-$	$sgn(k-K)$	$+$	$+\left(-\right)$ *for* $k+k'>\left(<\right)\tilde{K}_t^p\left(k_-\right)$
i^{DW}	$+$	$+$	$sgn(K-k)$	$-$	$+\left(-\right)$ *for* $k+k'<\left(>\right)\tilde{K}_t^p\left(k_+\right)$
K	$-$	$-$	$+\left(-\right)$ *for large (small)* k	$+$	$+\left(-\right)$ *for* $k+k'>\left(<\right)\tilde{K}_t^p\left(\zeta_t^p\right)$
κ	$-$	$-$	$sgn(k-K)$	$+$	$+\left(-\right)$ *for* $k+k'<\left(>\right)2K$
λ_0	$-$	0	$sgn(K-k)$	$-$	$+\left(-\right)$ *for* $k+k'>\left(<\right)2K$
λ	$+$	$-$	$sgn(K-k)$	$-$	$+\left(-\right)$ *for* $k+k'>\left(<\right)2K$

where

$$\tilde{K}_t^p\left(k^w\right)=2K\cdot\frac{k^w-\tilde{M}^p\left(T-\tau_1^p-\left(t-\tau_1^p\right)^+\right)\exp\left(rT\right)-1\left\{t<\tau_1^p\right\}e^{r\left(T-\tau_1^p\right)}\frac{\left(\lambda-\lambda_0\right)\eta^p}{2}\frac{\partial\tau_1^p}{\partial A_2}}{1-\tilde{M}^p\left(T-\tau_1^p-\left(t-\tau_1^p\right)^+\right)\exp\left(rT\right)-1\left\{t<\tau_1^p\right\}e^{r\left(T-\tau_1^p\right)}\frac{\left(\lambda-\lambda_0\right)\eta^p}{2}\frac{\partial\tau_1^p}{\partial A_2}},$$

and

$$\tilde{M}^p\left(u\right)=\frac{\lambda}{2}\frac{\partial\tau_1^p\left(\eta^p;A_2,T\right)}{\partial A_2}\int_{T-u}^{T}e^{-rs}\frac{H_s^p\left(2\kappa+H_s^p\right)}{\left(\kappa+H_s^p\right)^2}\left(-\dot{H}_s^p\right)ds.$$

2.5.5 Model extensions

Our closed-form model has focused on homogeneous banks except initial reserve balance so far. However, it allows for a set of extensions, in which we are still able to get closed-form solutions and conduct comparative statics analysis. In the appendix, we introduce four pieces of extensions separately to discuss the effects of other Federal funds market factors on the trade dynamics. Our main extension is a heterogeneous-agent model, where we add peripheral traders, e.g. government-sponsored enterprises and other financial institutions without Fed Reserve accounts, to the existing group of banks. We assume the peripheral traders contact banks at a constant search intensity, and obtain closed-form solutions. Instead of conducting comparative statics, we estimate this extended model via simulated method of moments and evaluate the quantitative importance of the disintermediation effect of unconventional monetary policy. Section 2.6 describes the model setup and presents the quantitative analysis, while Appendix 2.D provides the derivations for the closed-form solutions.

We also provide other extensions in the appendix. Appendix 2.E introduces Federal funds brokerage to the market to study how the unconventional monetary policies affect the size of brokerage. We assume the brokers compete for matchmaking services via free entry with non-zero entry cost. Thus the size of brokerage is endogenously determined. In particular, IOER has disintermediation effect on brokerage by lowering the equilibrium size of active brokers in the market. Appendix 2.F considers the effects of payment shocks on the market trade dynamics. We introduce both lumpy and continuous shocks to payment flows. In particular, we find that the payment shocks do not impact the equilibrium length of search and bilateral transaction size. Appendix 2.G discusses the effects of counterparty risk on the Federal funds trade. By counterparty risk, we assume both counterparties of a meeting could default on the trade independently with some constant probabilities. We find that the effects of higher counterparty risk are isomorphic to the effects of higher transaction costs or lower search intensity.

2.6 Quantitative analysis

This section provides a quantitative evaluation for the effects of unconventional monetary policy on disintermediation. The evaluation is based on an extended model that captures the main institutional features of the Federal funds market. The setup is as follows. There are two groups of agents: a unit continuum of banks as in the baseline model, and a continuum of peripheral traders that have no Federal reserve accounts. The peripheral traders represent government-sponsored enterprises and other financial institutions that participate in the Federal funds market but have no access to IOER. The mass of peripheral traders is ϑ. We assume a peripheral trader only contacts banks at a constant arrival rate φ. Moreover, the banks choose search intensity ε in the contact with other banks, at an arrival rate $m(\varepsilon, \varepsilon')$. The bargaining power of banks in the meeting with peripheral traders is $\theta \in (0,1)$. Each peripheral trader is endowed with some reserve balances $\tilde{k}$, and we denote the distribution of peripheral traders' reserve balances as $\tilde{F}_t\left(\tilde{k}\right)$, with $\tilde{F}_0\left(\tilde{k}\right)$ given.[15] We assume the peripheral traders have no flow payoff of reserve holdings, but only enjoy the end-of-period payoff from the overnigh reverse repurchase facility (ON RRP), i.e. $\tilde{U}\left(\tilde{k}\right) = \left(1 + i^{RRP}\right)\tilde{k}$.

For quantitative motivation, we assume the transaction cost of a bank in a meeting is $\chi(\varepsilon, q) = (\kappa_0 + \kappa_1 \varepsilon) q^2$. The peripheral traders are not subject to balance sheet regulations, thus their transaction cost is assumed to be 0. Since banks do not choose search intensity in contacting peripheral traders, their transaction costs in such contacts is $\kappa_0 q^2$. This extended model has closed-form solutions and Appendix 2.D presents the derivations. In particular, we find the banks' value functions are still quadratic and the peripheral traders' value functions are linear in their reserve balances.

To capture the change in the regulatory requirement on bank balance sheet and the opportunity cost of liquidity, we allow for time-varying transaction cost and liquidity benefits.

[15] As is shown in Appendix 2.D, the distribution $\tilde{F}_t\left(\tilde{k}\right)$ is redundant in equilibrium.

Specifically, we assume κ_0 and γ change over years in the following form:

$$\begin{aligned}\kappa_{0,yr} &= \kappa_{0,2006} \times \exp\left[g_{\kappa_0}\left(yr - 2006\right)\right], \\ \gamma_{yr} &= \gamma_{2006} \times \exp\left[g_{\gamma}\left(yr - 2006\right)\right],\end{aligned}$$

where yr denotes a year and takes values from 2006 to 2018. In our estimation, we set 2006 as the first year and 2018 as the last year of the sample. Therefore, instead of estimating g_{κ_0} and g_γ, we estimate $\kappa_{0,2018}$ and γ_{2018}.

2.6.1 Estimation

Instead of calibrating the deterministic theoretical model, we conduct a simulated method of moments estimation on a discretized version of the model to pin down the parameters. In the discretized version, we assume the reserve distribution is atomic (so there is a finite number of banks) and given by the empirical distribution of reserve balances in the data. The outcome of the discretized model is random since each bank faces idiosyncratic random meetings. We estimate the model parameters via simulated method of moments. The Appendix 2.H describes the algorithm of simulation and estimation.

In the current version of estimation, we first normalize $r = a_1 = a_2 = 0$, and set $T = 2.5/24$ to represent the 2.5 hr trading session of the daily Federal funds market. Second, we normalize the size of peripheral traders $\vartheta = 1$ since it cannot be identified separately from the contact rate φ. Third, the individual excess reserves are the quarterly bank-level data (Call reports and Form FR Y9-C) of individual excess reserves before Federal funds trade divided by bank assets. The data of IOER, primary credit rate and ON RRP are obtained from FRED. We conduct the simulated method of moments based on the data over 2006Q1-2018Q4 to estimate the following parameters

$$\left\{\lambda, \lambda_0, k_+, k_-, \gamma_{2006}, \gamma_{2018}, \theta, \varphi, \kappa_1, \kappa_{0,2006}, \kappa_{0,2018}\right\},$$

and the moments for estimation are (1) the regression coefficients of $i^{ER} \times k$ and $K \times k$ in the Federal funds net purchase regressions 2.2. (2) the banks' aggregate share of intermediation volume in 2006 and 2018; (3) the aggregate Fed funds sold by intermediaries normalized by aggregate bank assets in 2006 and 2018; (4) the aggregate Fed funds purchased by intermediaries normalized by aggregate bank assets in 2006 and 2018; (5) the aggregate fraction of trading banks in 2006 and 2018; (6) the average effective Fed funds rates in 2006 and 2018. The parameter estimation results are listed in Table 2.6. The simulated moments are listed in Table 2.7 and 2.8.

We find that the estimated transaction cost κ_0 increases from 2006 to 2008, while the liquidity benefit γ decreases in the same period. This implies the rise of bank balance sheet cost due to stronger regulations, and the declined liquidity benefit due to the increasing aggregate excess reserves. The moments produced by our estimation are close to the targets. In particular, the simulated regression coefficients have the correct signs and similar magnitudes, and the fraction of trading banks and effective Federal funds rates are almost exactly calibrated.

2.6.2 Counterfactual analysis

Given the estimation we conduct counterfactual analysis to evaluate the quantitative importance of unconventional monetary policies and regulations to the disintermediation channel. In particular, we consider the following exercises and examine how the level of intermediation in 2018 changes: (1) Change the paths of IOER, primary credit rate and ON RRP in 2018 to the paths in 2006. This exercise investigates how the level of intermediation changes in 2018 if the Federal Reserve recovers the policy rates in 2006. (2) Proportionally change individual banks' reserve balances in 2018, such that the average individual reserve balances are equal to the levels in 2006. This exercise examines the effect of aggregate excess reserves on disintermediation. (3) Change $\kappa_{0,2018}$ to $\kappa_{0,2006}$. This exercise evaluates the impact of rising transaction cost on disintermediation.

Table 2.9 reports the results of counterfactuals. We find that eliminating IOER doubles the intermediation volume share in 2018, while reducing the transaction cost can increase the level of intermediation by about 4 times. However, the effect of aggregate excess reserves on disintermediation is small, since the intermediation share almost doesn't change in the counterfactual analysis.

2.7 Conclusion

This paper proposes a new channel of monetary policy and regulation on the monetary policy implementation, the disintermediation channel. When the interest rate on excess reserves (IOER) increases or the balance sheet cost rises, the intermediation trades by banks decline in the Federal funds market. We rationalize this channel in a continuous-time search-and-bargaining model of divisible funds and endogenous search intensity, which nests the matching model of Afonso and Lagos (2015b) and the transaction model of Hamilton (1996). IOER decreases the spread of marginal value of reserves, and balance sheet cost increases the marginal cost of holding reserves, both of which lower the gains of intermediation. We find that the equilibrium is constrained inefficient as banks trade too frequently. The disintermediation channel is both empirically and quantitatively important. Empirically, it significantly impede the reallocation of reserves from lender banks to borrower banks. Quantitatively, eliminating IOER and reducing the balance sheet cost can greatly raise the level of intermediation during the period after the Great Recession. For further research, we will focus on the investigating how the disintermediation channel impacts the effects of current monetary policy framework on the Federal funds rate and real economy, as well as calculating the optimal monetary policy and regulation via quantitative analysis.

2.A Appendix: Details of data and measurement

In this section, we describe how we collect the data and construct various measurement we used for the summary statistics and estimation.

2.A.1 Sources

Financial data of the Federal funds market participants come from the following:

- **Call Reports**. This is the source of the subsidiary-level data. In particular, we use form FFIEC 031 for banks with both domestic and foreign offices, form FFIEC 041 for banks with domestic offices only, and form FFIEC 002 for U.S. branches and agencies of foreign banks (FBO). These forms are available for download at the Federal Financial Institutions Examination Council (FFIEC).[16]

- **FR Y-9C**. This is the source of the consolidated data at the level of holding companies (for bank holding companies, savings and loan holding companies, and intermediate holding companies) with total consolidated assets of $1 billion or more (prior to 2015, this threshold was just $500 million). This is available for download at the Federal Reserve Bank of Chicago.[17]

- **Attributes, relationships, and transformations tables**. This is the source of the ownership structure of holding companies upon their subsidiaries. They are available for download at National Information Center (NIC).[18]

- **10Q and 10K**. This is the source of government sponsored enterprises (GSE) data. These forms are available for download at the Security Examination Commission

[16] `https://cdr.ffiec.gov/public/`

[17] `https://www.chicagofed.org/banking/financial-institution-reports/bhc-data`

[18] `https://www.ffiec.gov/npw/FinancialReport/DataDownload`

(SEC).[19] The GSE data is fully available since 2006Q1.

- **H.4.1**. This is the source of the balance sheet of the Federal Reserve System and factors affecting reserve balances of depository institutions. This is available for download at the Board of Governors of the Federal Reserve System.[20]

- **Time series of the economy**. It is available for download at the Federal Reserve Bank of St Louis (FRED).[21]

2.A.2 Consolidated sample

Whenever possible, we always measure variables at the holding-company level. We think that holding companies are desirable sample unit because first, usually the subsidiaries' reserves, which are not directly observable in the Call reports, are corresponded by their holding company's master accounts in the Federal Reserve Banks, which are observable. Second, sometimes the decision of Federal Funds trading is delegated to the holding company. Third, it avoids double-counting the intra-holding-company Federal Funds trades, which are different from those normal interbank transactions.

Consolidation is done by referring to items filed in FR Y-9C. For the holding companies not eligible to file FR Y-9C, or items not available from FR Y-9C, we directly consolidate the Call report items from the subsidiary level up to the topmost holding-company level, based on the relationships table from NIC. In this appendix, we always refer i as the index for holding companies and j as the index for i's subsidiaries. We focus on banks that have positive amounts of asset and total reserve balances, and trade at least once in the Federal funds market in the data sample.

[19] `https://www.sec.gov/edgar/searchedgar/companysearch.html`

[20] `https://www.federalreserve.gov/releases/h41/`

[21] `https://fred.stlouisfed.org/`

2.A.3 Excess reserves

The formula to measure excess reserves bank i holds at the Federal Reserve account at the end of quarter t is given by

$$Excess\ Reserves_{it} = Total\ Reserves_{it} - \left\{\sum\nolimits_j Required\ Reserves_{jt} - Vault\ Cash_{it}\right\}_+ .$$

$Total\ Reserves_{it}$ is measured by item RCFD0090 in FR Y-9C ("Balances due from Federal Reserve Banks"). $Vault\ Cash_{it}$ is approximated by item RCON0080 in FR Y-9C ("Currency and coin"). The formula to calculate $Required\ Reserves_{jt}$ is based on subsidiary j's net transaction accounts. For example, the formula of reserve requirement in 2010 is given by the following table:

Table 2.1: Reserve requirement in 2010

Net transaction accounts	% required
\$0 to \$10.7 million	0
More than \$0.7 million to \$55.2 million	3
More than \$55.2 million	10

The table is updated every year.[22] To estimate net transaction acccounts, we substract item RCON 2215 of j's Call Report ("Total Transaction Accounts") from the sum of item RCFD 0083 ("Balances due from depository institutions in the U.S.: U.S. branches and agencies of foreign banks (including their IBFs)"), item RCFD 0085 ("Balances due from depository institutions in the U.S.: Other depository institutions in the U.S. (including their IBFs)") and item RCON 0020 ("Cash items in process of collection and unposted debit"). Then we apply the historical reserve requirement formulas on net trans accounts to calculate $Required\ Reserves_{jt}$.

[22]The historical reserve requirement can be found on `https://www.federalreserve.gov/monetarypolicy/reservereq.htm`

To measure the excess reserves bank i holds before entering the Federal funds market, we subtract the net Federal funds purchase from *Excess Reserves*$_{it}$. Thus the pre-trade excess reserves is given by

$$\begin{aligned} \textit{Excess Reserves pre-trade}_{it} \;=\; & \textit{Excess Reserves}_{it} \text{ - } \textit{Federal funds purchased}_{it} \\ & + \textit{ Federal funds sold}_{it}. \end{aligned}$$

By dividing the pre-trade excess reserves by bank assets, we obtain the measure $exres\_assets$ in the regressions.

2.A.4 Federal funds trades and intermediation

We compute the net Federal funds borrowed by substracting item BHDM B993 in FR Y-9C ("Federal funds purchased in domestic offices") from item BHDM B987 ("Federal funds sold in domestic offices"). We measure bank's intermediation by *Reallocated Funds*$_{it}$:

$$\begin{aligned} \textit{Reallocated Funds}_{it} \;=\; & \textit{Federal funds purchased}_{it} + \textit{Federal funds sold}_{it} \\ & - \left| \textit{Federal funds purchased}_{it} \text{ - } \textit{Federal funds sold}_{it} \right|. \end{aligned}$$

By dividing the net Federal funds borrowed and *Reallocated Funds* by bank assets respectively, we obtain the measure $ffnet\_assets$ and $ffreallo\_assets$ in the regressions.

2.A.5 Bank-level controls

We use the following items from Call report to measure various attributes of banks.

- Size and scope
  - logarithm of assets (item RCFD 2170 "Total assets").
  - bank equity (item RCFD 3210 "Total bank equity capital") over bank assets.

- Marginal benefit of liquidity
  - ROA
  - High-quality liquid assets (HQLA) over total assets (Ihrig et al., 2019)
- Risk
  - ratio of non-performing loan (sum of items 1 through 8.b of Column B and C in Schedule RC-N) over bank assets, as in Afonso et al. (2011)
  - ratio of loan (item RCFD 2122 "Total loans and leases held for investment and held for sale") over bank assets
- Regulation
  - Tier-1 leverage ratio (item RCFA 7204 "Tier 1 leverage ratio")
- Other indicators
  - bank entity type (in the NIC attributes table)
  - Fed District dummy (in the NIC attributes table)

2.A.6 Economy-wide controls

- quarterly real GDP growth rate (available from FRED)
- quarterly unemployment rate (available from FRED)
- standard deviation of the Fed's general treasury account in a quarter (available from H.4.1)

2.B Appendix: Tables

2.B.1 Summary statistics

Table 2.2: Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Net Fed funds purchase/Assets	107,959	-0.0074	0.0554	-0.9690	0.9608
Ex. res. pre-trade/Assets	107,959	0.0434	0.1015	-0.9062	4.1827
log (Assets)	107,959	13.7981	1.4571	4.6728	21.6874
Dummy: reallocation	52,778	0.2174	0.4125	0	1
Fed funds reallocation/Assets	52,778	0.0021	0.0072	0	0.0388
IOER (%)	64	0.3602	0.5470	0	2.4
Primary credit rate (%)	64	2.0781	0.1804	0.5	6.25
Agg. ex. res./Agg. assets	64	0.0428	0.0410	-0.0084	0.1070

Notes: This table presents the summary statistics of key variables. The observations for the first 5 variables are bank-quarter. "Net Fed funds purchase/Assets" is a bank's net Federal funds purchase divided by bank assets. "Ex. res. pre-trade/Assets" is a bank's excess reserve balances before Federal funds trade divided by bank assets. "log(Assets)" is the log value of bank assets. "Dummy: reallocation" is equal to 1 if a bank intermediates Federal funds on a day, and equal to 0 otherwise. "Fed funds reallocation/Assets" is a bank's volume of Federal funds reallocation divided by bank assets. "Agg. ex. res/Agg. assets" is the aggregate excess reserve balances before Federal funds trade divided by the aggregate bank assets. The sample consists of U.S. banks that hold positive total reserves at the Fed account and trade Federal funds at least once in the data sample. The sample period is from 2003Q1 to 2018Q4.

2.B.2 Regression results

Table 2.3: Probit on Reallocation

Dep. Var.	Dummy: Reallocation					
	Probit (Pooled)		Panel Probit (RE)		IV Probit	
	(1)	(2)	(3)	(4)	(5)	(6)
IOER	-0.182***	-0.179***	-0.320***	-0.318***	-0.639***	-0.721***
	(0.033)	(0.037)	(0.051)	(0.055)	(0.085)	(0.100)
IOER×Ind. ex res		1.250		0.397		5.352***
		(0.986)		(1.162)		(1.450)
Agg ex res	-3.400***	-4.312***	-5.871***	-5.843***	-3.183***	-3.686***
	(0.616)	(0.759)	(0.935)	(1.006)	(0.752)	(0.969)
Agg ex res×Ind. ex res		46.293**		-2.055		23.063
		(17.992)		(21.176)		(24.490)
Prim. credit rate	-0.003	-0.002	-0.025	-0.021	-0.050***	-0.051***
	(0.011)	(0.012)	(0.016)	(0.017)	(0.019)	(0.019)
Prim. credit rate		-0.480*		-0.625		-0.634
×Ind. ex res		(0.260)		(0.388)		(0.485)
Ind. ex res	-4.596***	-6.275***	-10.270***	-9.046***	-3.426***	-5.256***
	(0.911)	(0.981)	(1.223)	(1.624)	(0.210)	(1.891)
All Fixed Effects	Y	Y	Y	Y	Y	Y
Bank controls	Y	Y	Y	Y	Y	Y
Agg. controls	Y	Y	Y	Y	Y	Y
Specification tests						
Wald test of exogeneity						
χ^2 stat					60.72	53.09
p-value					[0.000]	[0.000]
Weak instrument test						
χ^2 stat					962.82	1093.15
p-value					[0.000]	[0.000]
Hansen J test						
χ^2 stat						2.188
p-value						[0.335]
Pseudo R^2	0.174	0.174				
Number of observations	44,048	44,048	44,048	44,048	39,674	39,674
Number of banks	1,122	1,122	1,122	1,122	1,121	1,121

Notes: This table presents the estimation results on the Probit regression of Federal funds intermediation (2.1). The sample consists of U.S. banks that hold positive total reserves at the Fed account and intermediate Federal funds at least once in the data sample. The sample period is from 2003Q1 to 2018Q4. Standard errors clustered by banks are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.4: Tobit on Reallocation

Dep. Var.	FF Reallocation/Assets					
	Tobit (Pooled)		Panel Tobit (RE)		IV Tobit	
	(1)	(2)	(3)	(4)	(5)	(6)
IOER	-0.002***	-0.002***	-0.002***	-0.002***	-0.009***	-0.010***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
IOER×Ind. ex res		0.008		-0.001		0.087***
		(0.021)		(0.004)		(0.020)
Agg ex res	-0.058***	-0.074***	-0.059***	-0.055***	-0.057***	-0.066***
	(0.009)	(0.012)	(0.007)	(0.007)	(0.011)	(0.013)
Agg ex res×Ind. ex res		0.846***		-0.185***		0.424
		(0.257)		(0.061)		(0.298)
Prim. credit rate	-0.000	-0.000	-0.000***	-0.000***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Prim. credit rate		-0.000		0.001		-0.001
×Ind. ex res		(0.003)		(0.001)		(0.005)
Ind. ex res	-0.070***	-0.097***	-0.042***	-0.040***	-0.058***	-0.086***
	(0.012)	(0.012)	(0.002)	(0.004)	(0.003)	(0.021)
All Fixed Effects	Y	Y	Y	Y	Y	Y
Bank controls	Y	Y	Y	Y	Y	Y
Agg. controls	Y	Y	Y	Y	Y	Y
Specification tests						
Wald test of exogeneity						
χ^2 stat					14.52	37.88
p-value					[0.006]	[0.000]
Weak instrument test						
χ^2 stat					1098.27	1468.42
p-value					[0.000]	[0.000]
Hansen J test						
χ^2 stat						5.439
p-value						[0.066]
Pseudo R^2	-0.298	-0.312				
Number of observations	44,097	44,097	44,097	44,097	39,691	39,691
Number of banks	1,127	1,127	1,127	1,127	1,122	1,122

Notes: This table presents the estimation results on the Tobit regression of Federal funds intermediation (2.1). The sample consists of U.S. banks that hold positive total reserves at the Fed account and intermediate Federal funds at least once in the data sample. The sample period is from 2003Q1 to 2018Q4. Standard errors clustered by banks are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.5: Effects of IOER and aggregate excess reserves on net Federal funds purchased

Dep. Var.	ffnet_assets					
	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
IOER	0.001	0.002***		-0.073***	-0.006	
	(0.001)	(0.001)		(0.011)	(0.024)	
IOER×Ind. ex res		0.102***	0.102***		0.300***	0.301***
		(0.011)	(0.011)		(0.062)	(0.069)
Agg ex res	-0.025***	-0.077***		-0.426***	-0.226*	
	(0.010)	(0.013)		(0.056)	(0.116)	
Agg ex res×Ind. ex res		2.893***	2.910***		4.221***	4.252***
		(0.288)	(0.291)		(0.834)	(0.830)
Prim. credit rate	-0.001***	0.000		-0.018***	-0.001	
	(0.000)	(0.000)		(0.002)	(0.005)	
Prim. credit rate		-0.071***	-0.071***		-0.059***	-0.059***
×Ind. ex res		(0.009)	(0.009)		(0.018)	(0.018)
Ind. ex res	-0.361***	-0.434***	-0.435***	-0.310***	-0.591***	-0.595***
	(0.021)	(0.037)	(0.037)	(0.0037)	(0.077)	(0.076)
Bank FE	Y	Y	Y	Y	Y	Y
Quarter FE	N	N	Y	N	N	Y
Year FE	Y	Y	N	Y	Y	N
Bank controls	Y	Y	Y	Y	Y	Y
Agg. controls	Y	Y	Y	Y	Y	Y
Specification tests						
Underidentification test						
χ^2 stat				291.4	251.8	73.8
p-value				[0.000]	[0.000]	[0.000]
Weak Instrument test						
F stat				70.31	31.68	10.08
10% relative bias (p-val)				[0.000]	[0.000]	[0.000]
30% relative bias (p-val)				[0.000]	[0.000]	[0.000]
Hansen J test						
χ^2 stat					0.0787	0.0497
p-value					[0.779]	[0.824]
Adj. R^2	0.810	0.865	0.866	0.305	0.504	0.498
Number of observations	104,291	104,291	104,291	85,141	85,141	85,141
Number of banks	3,506	3,506	3,506	2,909	2,909	2,909

Notes: This table presents the estimation results on the net Federal funds purchased regression (2.2). The sample consists of U.S. banks that hold positive total reserves at the Fed account and trade Federal funds at least once in the data sample. The sample period is from 2003Q1 to 2018Q4. The fixed effects of Bank×Crisis and Bank×Post-Crisis are also controlled. Standard errors clustered by banks are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

2.B.3 Tables in quantitative analysis

Table 2.6: Parameter estimation

Parameter	λ	λ_0/λ_0	k_+	k_-	θ	ρ
Estimated Value	20.1987	0.5605	2.9480	-0.0596	0.7005	0.2000
Standard deviation	0.0007	3.6×10^{-5}	0.0038	0.0048	0.0043	1.3×10^{-5}
Parameter	κ_1	$\kappa_{0,2006}$	$\kappa_{0,2018}$	γ_{2006}	γ_{2018}	
Estimated Value	0.00568	0.00001	0.000705	0.00035	0.00028	
Standard deviation	0.0054	0.0038	0.0024	0.0062	0.0019	

Notes: This table lists the estimated values and standard deviations of the model parameters from simulated method of moments.

Table 2.7: Simulated regression coefficients

Moments	Target	Simulation	95% CI
Coef of ind. ex. res.	-0.595	-0.199	[-0.252,-0.151]
Coef of ind. ex. res×ioer†	0.301	0.0540	[0.044,0.067]
Coef of ind. ex. res×dw	-0.059	-0.0084	[-0.015,-0.003]
Coef of ind. ex. res×agg. ex. res.†	4.252	2.2829	[1.548,3.069]

Notes: This table presents the simulated coefficients of Federal funds net purchase regressions under the estimated parameters. The column "Target" lists the estimated coefficients from the original regressions. The column "Simulation" lists the simulated coefficients. The column "95% CI" lists the 95% confidence interval of the simulated coefficients. The sign † represents the target is used in estimation. "ind. ex. res." is the individual excess reserves divided by individual bank assets. "ioer" is the interest rate on excess reserves. "dw" is the primary credit rate. "agg. ex. res." is the aggregate excess reserves divided by aggregate bank assets.

Table 2.8: Simulated moments

Year	2006		2018	
	Target	Simulation	Target	Simulation
Intermediation volume share	0.2150	0.1715	0.0663	0.0726
FF sold by intermediary	0.0045	0.0034	0.0002	0.0009
FF purchased by intermediary	0.0107	0.0062	0.0013	0.0031
Fraction of trading banks	0.8894	0.8805	0.6896	0.6985
Effective Federal funds rate	0.0514	0.0511	0.0204	0.0207

Notes: This table presents the simulated moments under the estimated parameters. The column "Target" lists the moments from the data. The column "Simulation" lists the simulated moments. All the targets are used in estimation. "Intermediation volume share" is the share of Federal funds reallocation in total Federal funds volume. "FF sold by intermediary" is the volume of Federal funds sold by intermediary banks as a share of aggregate bank assets. "FF purchased by intermediary" is the volume of Federal funds purchased by intermediary banks as a share of aggregate bank assets. "Fraction of trading banks" is the fraction of banks that trade in the total number of banks. All the moments are average values across quarters within each year.

Table 2.9: Counterfactual analysis

			Counterfactual analysis		
			(1)	(2)	(3)
Moments in 2018	Target	Simulation	IOER	Agg ex res	Transct cost
Intermediation Volume Share	0.0663	0.0726	0.1328	0.0654	0.3025
FF sold by intermediary	0.0002	0.0009	0.0019	0.0006	0.0166
FF purchased by intermediary	0.0013	0.0031	0.0041	0.0029	0.0382
Fraction of trading banks	0.6896	0.6985	0.8802	0.6985	0.6985
Effective Federal funds rate	0.0204	0.0207	0.0318	0.0203	0.0331

Notes: This table presents the simulated counterfactual analysis under the estimated parameters. The column "Target" lists the moments from the data. The column "Simulation" lists the simulated moments of the estimated model. The columns under "Counterfactual analysis" lists the simulated moments. under the corresponding counterfactual exercise. "IOER" represents the exercise that changes the values of IOER, primary credit rate and ON RRP from 2018 to 2006. "Agg ex res" represents the exercise that changes the aggregate excess reserves from 2018 to 2006 by proportionaly scaling individual excess reserves. "Transct cost" represents the exercise that changes the transaction parameter κ_0 from the 2018 value to 2006 value. "Intermediation volume share" is the share of Federal funds reallocation in total Federal funds volume. "FF sold by intermediary" is the volume of Federal funds sold by intermediary banks as a share of aggregate bank assets. "FF purchased by intermediary" is the volume of Federal funds purchased by intermediary banks as a share of aggregate bank assets. "Fraction of trading banks" is the fraction of banks that trade in the total number of banks. All the moments are average values across quarters within each year.

2.C Appendix: Proofs and derivations

2.C.1 Derivation of the general form of $m(\varepsilon, \varepsilon')$

For any $\varepsilon, \varepsilon' \in [0,1]$, equation (2.4) implies that

$$\begin{aligned} m(\varepsilon, \varepsilon') &= \varepsilon' m(\varepsilon, 1) + (1-\varepsilon') m(\varepsilon, 0) \\ &= [m(\varepsilon, 1) - m(\varepsilon, 0)]\varepsilon' + m(\varepsilon, 0). \end{aligned}$$

By symmetry we have

$$\begin{aligned} m(\varepsilon, 1) &= m(1, \varepsilon) = [m(1,1) - m(1,0)]\varepsilon + m(1,0), \\ m(\varepsilon, 0) &= m(0, \varepsilon) = [m(0,1) - m(0,0)]\varepsilon + m(0,0). \end{aligned}$$

Thus we can get

$$\begin{aligned} m(\varepsilon, \varepsilon') &= [m(\varepsilon,1) - m(\varepsilon,0)]\varepsilon' + m(\varepsilon,0) \\ &= \{[m(1,1) - m(1,0)]\varepsilon + m(1,0) - [m(0,1) - m(0,0)]\varepsilon - m(0,0)\}\varepsilon' \\ &\quad + [m(0,1) - m(0,0)]\varepsilon + m(0,0) \\ &= [m(1,1) - m(1,0) - m(0,1) + m(0,0)]\varepsilon\varepsilon' + [m(0,1) - m(0,0)]\varepsilon \\ &\quad + [m(1,0) - m(0,0)]\varepsilon' + m(0,0) \\ &= (\lambda - 2\lambda_1 + \lambda_0)\varepsilon\varepsilon' + (\lambda_1 - \lambda_0)(\varepsilon + \varepsilon') + \lambda_0. \end{aligned}$$

2.C.2 Proof of Lemma 2.1

Proof. (i). [S_t decreases in ε and ε']: Pick any $\varepsilon, \tilde{\varepsilon}$ s.t. $\tilde{\varepsilon} > \varepsilon$,

$$\begin{aligned}
S_t(k, k', \varepsilon, \varepsilon') &= V_t[k + q_t(k, k', \varepsilon, \varepsilon')] - V_t(k) - \chi[\varepsilon, q_t(k, k', \varepsilon, \varepsilon')] \\
&\quad + V_t[k' - q_t(k, k', \varepsilon, \varepsilon')] - V_t(k') - \chi[\varepsilon', -q_t(k, k', \varepsilon, \varepsilon')] \\
&\geq V_t[k + q_t(k, k', \tilde{\varepsilon}, \varepsilon')] - V_t(k) - \chi[\varepsilon, q_t(k, k', \tilde{\varepsilon}, \varepsilon')] \\
&\quad + V_t[k' - q_t(k, k', \tilde{\varepsilon}, \varepsilon')] - V_t(k') - \chi[\varepsilon', -q_t(k, k', \tilde{\varepsilon}, \varepsilon')] \\
&\geq V_t[k + q_t(k, k', \tilde{\varepsilon}, \varepsilon')] - V_t(k) - \chi[\tilde{\varepsilon}, q_t(k, k', \tilde{\varepsilon}, \varepsilon')] \\
&\quad + V_t[k' - q_t(k, k', \tilde{\varepsilon}, \varepsilon')] - V_t(k') - \chi[\varepsilon', -q_t(k, k', \tilde{\varepsilon}, \varepsilon')] \\
&= S_t(k, k', \tilde{\varepsilon}, \varepsilon').
\end{aligned}$$

Since S_t is symmetric in ε and ε', then S_t also decreases in ε'.

[$|q_t|$ decreases in ε and ε']: Since $\chi(\varepsilon, q)$ is complementary in ε and q, then for any ε', ε, q', q such that $\varepsilon' > \varepsilon$ and $q' > q \geq 0$, we have

$$\chi(\varepsilon', q') - \chi(\varepsilon', q) \geq \chi(\varepsilon, q') - \chi(\varepsilon, q).$$

This means the function $h(\varepsilon; q', q) := \chi(\varepsilon, q') - \chi(\varepsilon, q)$ is a single crossing function for any ε and $q' > q \geq 0$. By Milgrom and Shannon (1994),

$$q_t(k, k', \varepsilon, \varepsilon') = \arg\max_q \{V_t(k + q) + V_t(k' - q) - \chi(\varepsilon, q) - \chi(\varepsilon', q)\} \tag{2.54}$$

is decreasing in ε if $q_t(k, k', \varepsilon, \varepsilon') > 0$, and increasing in ε if $q_t(k, k', \varepsilon, \varepsilon') < 0$. Similarly, we can prove $|q_t(k, k', \varepsilon, \varepsilon')|$ is decreasing in ε'.

[S_t supermodular]: Suppose $V_t(k)$ is weakly concave and twice differentiable, then the optimal trade size $q_t(k, k', \varepsilon, \varepsilon')$ is interior and differentiable by the implicit function theorem. Without loss of generality we assume $q_t(k, k', \varepsilon, \varepsilon') > 0$. Then by the envelope theorem we

have

$$\frac{\partial S_t\left(k,k',\varepsilon,\varepsilon'\right)}{\partial\varepsilon}=-\chi_\varepsilon\left(\varepsilon,q_t\left(k,k',\varepsilon,\varepsilon'\right)\right),$$

and

$$\frac{\partial^2 S_t\left(k,k',\varepsilon,\varepsilon'\right)}{\partial\varepsilon\partial\varepsilon'}=-\chi_{\varepsilon q}\left(\varepsilon,q_t\left(k,k',\varepsilon,\varepsilon'\right)\right)\frac{\partial q_t\left(k,k',\varepsilon,\varepsilon'\right)}{\partial\varepsilon'}>0,$$

where we apply $\chi_{\varepsilon q}>0$ and $\frac{\partial q_t(k,k',\varepsilon,\varepsilon')}{\partial\varepsilon'}<0$.

(ii). $[S_t\left(k,k,\varepsilon,\varepsilon'\right)=0]$ If $V_t\left(k\right)$ is concave, then

$$\begin{aligned}
S_t\left(k,k,\varepsilon,\varepsilon'\right) &\equiv \max_q\left\{V_t\left(k+q\right)+V_t\left(k-q\right)-V_t\left(k\right)-V_t\left(k\right)-\chi\left(\varepsilon,q\right)-\chi\left(\varepsilon',-q\right)\right\}\\
&\leq \max_q\left\{2V_t\left(k\right)-V_t\left(k\right)-V_t\left(k\right)-\chi\left(\varepsilon,q\right)-\chi\left(\varepsilon',-q\right)\right\}\\
&= \max_q\left\{-\chi\left(\varepsilon,q\right)-\chi\left(\varepsilon',q\right)\right\}=0.
\end{aligned}$$

[Monotonicity of $S_t\left(k,k',\varepsilon,\varepsilon'\right)$ and $q_t\left(k,k',\varepsilon,\varepsilon'\right)$ in k] Note that for a concave functio $f\left(x\right)$, where x is a scalar, we must have that for any $x'>x$ and $\Delta>0$,

$$\begin{aligned}
f\left(x\right)+f\left(x'\right) &= f\left(\frac{x'-x+\Delta}{x'-x+2\Delta}\left(x-\Delta\right)+\frac{\Delta}{x'-x+2\Delta}\left(x'+\Delta\right)\right)\\
&\quad +f\left(\frac{\Delta}{x'-x+2\Delta}\left(x-\Delta\right)+\frac{x'-x+\Delta}{x'-x+2\Delta}\left(x'-\Delta\right)\right)\\
&\geq \frac{x'-x+\Delta}{x'-x+2\Delta}f\left(x-\Delta\right)+\frac{\Delta}{x'-x+2\Delta}f\left(x'+\Delta\right)\\
&\quad +\frac{\Delta}{x'-x+2\Delta}f\left(x-\Delta\right)+\frac{x'-x+\Delta}{x'-x+2\Delta}f\left(x'+\Delta\right)\\
&= f\left(x-\Delta\right)+f\left(x'+\Delta\right).
\end{aligned}\tag{2.55}$$

Thus for any $k'>k$ and $q<0$:

$$\begin{aligned}
&V_t\left(k+q\right)+V_t\left(k'-q\right)-\chi\left(\varepsilon,q\right)-\chi\left(\varepsilon',-q\right)\\
<\ &V_t\left(k\right)+V_t\left(k'\right)-\chi\left(\varepsilon,0\right)-\chi\left(\varepsilon',0\right),
\end{aligned}$$

which implies that $q_t(k, k', \varepsilon, \varepsilon') \geq 0$ for any $k' > k$, with strict inequality if V_t is strictly concave. Moreover, for any $\tilde{k} > k$ and $\tilde{q} > q$, the inequality (2.55) implies that

$$\begin{aligned} & V_t(k+\tilde{q}) + V_t\left(\tilde{k}+q\right) \\ \geq\ & V_t(k+\tilde{q}-(\tilde{q}-q)) + V_t\left(\tilde{k}+q+(\tilde{q}-q)\right) \\ =\ & V_t(k+q) + V_t\left(\tilde{k}+\tilde{q}\right). \end{aligned}$$

Therefore, the function $V_t(k+q)$ has increasing differences over $(-k, q)$. This implies that for any $k' > \tilde{k} > k$, and any ε and ε',

$$\begin{aligned} & S_t(k, k', \varepsilon, \varepsilon') \\ =\ & V_t(k + q_t(k, k', \varepsilon, \varepsilon')) + V_t(k' - q_t(k, k', \varepsilon, \varepsilon')) - V_t(k) - V_t(k') \\ & -\chi(\varepsilon, q_t(k, k', \varepsilon, \varepsilon')) - \chi(\varepsilon', -q_t(k, k', \varepsilon, \varepsilon')) \\ \geq\ & V_t\left(k + q_t\left(\tilde{k}, k', \varepsilon, \varepsilon'\right)\right) - V_t(k) + V_t\left(k' - q_t\left(\tilde{k}, k', \varepsilon, \varepsilon'\right)\right) - V_t(k') \\ & -\chi\left(\varepsilon, q_t\left(\tilde{k}, k', \varepsilon, \varepsilon'\right)\right) - \chi\left(\varepsilon', -q_t\left(\tilde{k}, k', \varepsilon, \varepsilon'\right)\right) \\ \geq\ & V_t\left(\tilde{k} + q_t\left(\tilde{k}, k', \varepsilon, \varepsilon'\right)\right) - V_t\left(\tilde{k}\right) + V_t\left(k' - q_t\left(\tilde{k}, k', \varepsilon, \varepsilon'\right)\right) - V_t(k') \\ & -\chi\left(\varepsilon, q_t\left(\tilde{k}, k', \varepsilon, \varepsilon'\right)\right) - \chi\left(\varepsilon', -q_t\left(\tilde{k}, k', \varepsilon, \varepsilon'\right)\right) \\ =\ & S_t\left(\tilde{k}, k', \varepsilon, \varepsilon'\right), \end{aligned}$$

where the inequality in the fourth line is due to the increasing differences property of $V_t(k+q)$ over $(-k, q)$. The inequality is strict if V_t is strictly concave. Similarly, we can prove $S_t(k, k', \varepsilon, \varepsilon')$ is (strictly) increasing in k for all $k > k'$. Moreover, by Milgrom and Shannon (1994), the increasing differences property also implies that for any $k' > k$, $q_t(k, k', \varepsilon, \varepsilon')$ is decreasing in k and increasing in k'. **Q.E.D.**

2.C.3 Derivation of HJB (2.8) and KFE (2.9)

By the property of Poisson process, the equation (2.7) for value function $V_t(k)$ can be rewritten as

$$V_t(k) = \max_{\{\varepsilon_z\}_{z\in[t,T]}\in[0,1]^{[t,T]}} \left\{ \begin{array}{l} \int_t^T e^{-\int_t^z [r+m(\varepsilon_s,\bar{\varepsilon}_s)]ds} \left\{ \begin{array}{l} u(k) + \\ \int_{k'} \left\{ \begin{array}{l} V_z\left[k+q_z\left(k,k',\varepsilon_z,\varepsilon_z\left(k'\right)\right)\right] \\ -\chi\left[\varepsilon_z, q_z\left(k,k',\varepsilon_z,\varepsilon_z\left(k'\right)\right)\right] \\ -e^{-r(T+\Delta-z)} R_z\left(k,k',\varepsilon_z,\varepsilon_z\left(k'\right)\right) \end{array} \right\} \\ \times m\left(\varepsilon_z,\varepsilon_z\left(k'\right)\right) dF_z\left(k'\right) \end{array} \right\} dz \\ +e^{-\int_t^T [r+m(\varepsilon_s,\bar{\varepsilon}_s)]ds} U(k) \end{array} \right\}$$

Denote $\varepsilon_t^*(k)$ as one equilibrium search profile. By taking the first-order derivative of $V_t(k)$ w.r.t. t and plugging in the solution to $e^{-r(T+\Delta-z)} R_z\left(k,k',\varepsilon_z,\varepsilon_z\left(k'\right)\right)$, we can obtain

$$rV_t(k) = \dot{V}_t(k) + u(k) + \int \frac{1}{2} S_t\left[k,k',\varepsilon_t^*(k),\varepsilon_t^*\left(k'\right)\right] m\left[\varepsilon_t^*(k),\varepsilon_t^*\left(k'\right)\right] dF_t\left(k'\right).$$

To derive the optimality condition for $\varepsilon_t^*(k)$, let $\mathbf{B}$ denote the space of bounded real-valued functions defined on $\mathbb{K}\times[0,T]$. Define a mapping $\mathcal{M}$ on $\mathbf{B}$ as follows:

$$(\mathcal{M}w)(k,t) = \max_{\{\varepsilon_z\}_{z\in[t,T]}\in[0,1]^{[t,T]}} \left\{ \begin{array}{l} \int_t^T e^{-\int_t^z [r+m(\varepsilon_s,\bar{\varepsilon}_s)]ds} \left\{ \begin{array}{l} u(k) + \\ \int_{k'} \left\{ \begin{array}{l} w\left[k+b_z\left(k,k',\varepsilon_z,\varepsilon_z\left(k'\right)\right),z\right] \\ -\chi\left[\varepsilon_z, b_z\left(k,k',\varepsilon_z,\varepsilon_z\left(k'\right)\right)\right] \\ -e^{-r(T+\Delta-z)} Y_z\left(k,k',\varepsilon_z,\varepsilon_z\left(k'\right)\right) \end{array} \right\} \\ \times m\left(\varepsilon_z,\varepsilon_z\left(k'\right)\right) dF_z\left(k'\right) \end{array} \right\} dz \\ +e^{-\int_t^T [r+m(\varepsilon_s,\bar{\varepsilon}_s)]ds} U(k) \end{array} \right\}$$

where

$$b_t(k,k',\varepsilon,\varepsilon\prime) \in \arg\max_b \left\{ \begin{array}{c} w(k+b,t) - w(k,t) - \chi(\varepsilon,b) \\ +w(k'-b,t) - w(k',t) - \chi(\varepsilon',-b) \end{array} \right\}$$

and

$$e^{-r(T+\Delta-t)} Y_t(k,k',\varepsilon,\varepsilon') = \frac{1}{2} \left\{ \begin{array}{c} w(k+b_t(k,k',\varepsilon,\varepsilon'),t) - w(k,t) - \chi(\varepsilon,b_t(k,k',\varepsilon,\varepsilon')) \\ +w(k',t) - w(k'-b_t(k,k',\varepsilon,\varepsilon'),t) + \chi(\varepsilon',-b_t(k,k',\varepsilon,\varepsilon')) \end{array} \right\}.$$

It is clear that the solution $V_t(k)$ to the HJB (2.8) is a fixed point of the mapping $\mathcal{M}$. Therefore, $\varepsilon_t^*(k)$ must be the solution to the right-hand side of $(\mathcal{M}w)(k,t)$ if we replace w with V. Note that since the time variable t is continuous, we have a continuum of control variables. We follow the heuristic approach in Van Imhoff (1982) to derive the condition for $\varepsilon_t^*(k)$. This approach relies on interpreting the integral in $(\mathcal{M}w)(k,t)$ as a summation of discrete variables over intervals with widths dz and dt. Then the Lebesgue dominated convergence theorem guarantees that the summation converges to the original integral as the widths of intervals approach 0. Then the terms in $(\mathcal{M}w)(k,t)$ which are related to $\varepsilon_t(k)$ can be written as

$$\begin{aligned} & e^{-\int_t^{t+dt}[r+m(\varepsilon_t(k),\bar{\varepsilon}_t)]ds} \left\{ \begin{array}{l} u(k) + \int_{k'} \left\{ \begin{array}{l} w[k+b_t(k,k',\varepsilon_t(k),\varepsilon_t(k')),t] \\ -\chi[\varepsilon_t(k),b_t(k,k',\varepsilon_t(k),\varepsilon_t(k'))] \\ -e^{-r(T+\Delta-t)} Y_t(k,k',\varepsilon_t(k),\varepsilon_t(k')) \end{array} \right\} \\ \times m(\varepsilon_t(k),\varepsilon_t(k'))\, dF_t(k') \end{array} \right\} dt \\ & + e^{-\int_t^{t+dt}[r+m(\varepsilon_t(k),\bar{\varepsilon}_t)]ds} w(k,t-dt) \\ = \; & (1-rdt)\, w(k,t-dt) + o(|dt|) + \left\{ \begin{array}{l} u(k) + \\ \int_{k'} \left\{ \begin{array}{l} w[k+b_t(k,k',\varepsilon_t(k),\varepsilon_t(k')),t] \\ -w(k,t-dt) \\ -\chi[\varepsilon_t,b_t(k,k',\varepsilon_t(k),\varepsilon_t(k'))] \\ -e^{-r(T+\Delta-t)} Y_t(k,k',\varepsilon_t(k),\varepsilon_t(k')) \end{array} \right\} \\ \times m(\varepsilon_t(k),\varepsilon_t(k'))\, dF_t(k') \end{array} \right\} dt. \end{aligned}$$

Thus the maximizer of $\varepsilon_t(k)$ to the above equation when $dt \to 0$ is given by

$$\varepsilon_t(k) \in \arg\max_{\varepsilon\in[0,1]} \left\{ \begin{array}{l} \int_{k'} \frac{1}{2} \left[\begin{array}{l} w\left[k + b_t\left(k,k',\varepsilon,\varepsilon_t\left(k'\right)\right),t\right] - w\left(k,t\right) \\ -\chi\left[\varepsilon_t, b_t\left(k,k',\varepsilon,\varepsilon_t\left(k'\right)\right)\right] \\ +w\left[k' - b_t\left(k,k',\varepsilon,\varepsilon_t\left(k'\right)\right),t\right] - w\left(k',t\right) \\ -\chi\left[\varepsilon_t\left(k'\right), -b_t\left(k,k',\varepsilon,\varepsilon_t\left(k'\right)\right)\right] \end{array} \right] \\ \times m\left(\varepsilon_t\left(k\right),\varepsilon_t\left(k'\right)\right) dF_t\left(k'\right) \end{array} \right\}$$

where we plug in the solution to $e^{-r(T+\Delta-t)}Y_t\left(k,k',\varepsilon,\varepsilon_t\left(k'\right)\right)$. This gives the HJB (2.8).

Next we take a heuristic approach to derive the KFE. Let Δ be a small time interval that is close to 0. Then by definition of $F_t(k)$, we have

$$\begin{aligned} F_{t+\Delta}\left(k^w\right) &= \left[1 - \Delta \cdot m\left(\varepsilon_t\left(k\right),\bar{\varepsilon}_t\right)\right] F_t\left(k^w\right) \\ &\quad + \int_{k\le k^w}\int_{k'} \Delta \cdot m\left(\varepsilon_t\left(k\right),\varepsilon_t\left(k'\right)\right) 1\left\{k + q_t\left(k,k'\right) \le k^w\right\} dF_t\left(k'\right) dF_t\left(k\right) \\ &\quad + \int_{k>k^w}\int_{k'} \Delta \cdot m\left(\varepsilon_t\left(k\right),\varepsilon_t\left(k'\right)\right) 1\left\{k + q_t\left(k,k'\right) \le k^w\right\} dF_t\left(k'\right) dF_t\left(k\right). \end{aligned}$$

On the right-hand side, the first term represents the mass of banks that do not meet counterparties during $[t, t+\Delta]$. The second term represents the banks that have meetings during $[t, t+\Delta]$ and hold reserves no more than k^w both before and after the meeting. The third term represents the banks that have meetings during $[t, t+\Delta]$ and hold reserves more than k^w before meeting and no more than k^w after the meeting. By rearranging terms, we can get

$$\begin{aligned} \frac{F_{t+\Delta}\left(k^w\right) - F_t\left(k^w\right)}{\Delta} &= -\int_{k\le k^w}\int_{k'} m\left(\varepsilon_t\left(k\right),\varepsilon_t\left(k'\right)\right) 1\left\{k + q_t\left(k,k'\right) > k^w\right\} dF_t\left(k'\right) dF_t\left(k\right) \\ &\quad + \int_{k>k^w}\int_{k'} m\left(\varepsilon_t\left(k\right),\varepsilon_t\left(k'\right)\right) 1\left\{k + q_t\left(k,k'\right) \le k^w\right\} dF_t\left(k'\right) dF_t\left(k\right), \end{aligned}$$

where we expand $m\left(\varepsilon_t\left(k\right),\bar{\varepsilon}_t\right)F_t\left(k^w\right)$ to

$$\int_{k\leq k^w}\int_{k'} m\left(\varepsilon_t\left(k\right),\varepsilon_t\left(k'\right)\right)dF_t\left(k'\right)dF_t\left(k\right),$$

and combine it with $\int_{k\leq k^w}\int_{k'} m\left(\varepsilon_t\left(k\right),\varepsilon_t\left(k'\right)\right)1\left\{k+q_t\left(k,k'\right)\leq k^w\right\}dF_t\left(k'\right)dF_t\left(k\right)$. Then we can take $\Delta\to 0$ and obtain the KFE (2.9).

2.C.4 Proof of Proposition 2.1

Proof. To prove $\left\{\Omega\left(S_t,F_t\right),\succeq_s\right\}$ is a complete lattice, it is sufficient to show $S_t\left(k,k',\varepsilon,\varepsilon'\right)m\left(\varepsilon,\varepsilon'\right)$ is supermodular in ε and ε'. Appendix 2.C.1 implies that $\theta_m\left(\varepsilon\right)=\frac{(\lambda-2\lambda_1+\lambda_0)\varepsilon}{(\lambda-2\lambda_1+\lambda_0)\varepsilon+\lambda_1-\lambda_0}$ only depends on ε, and

$$m_1\left(\varepsilon,\varepsilon'\right)=m_2\left(\varepsilon',\varepsilon\right)=\frac{m_{12}\left(\varepsilon',\varepsilon\right)\varepsilon'}{\theta_m\left(\varepsilon'\right)}=\frac{m_{12}\left(\varepsilon,\varepsilon'\right)\varepsilon'}{\theta_m\left(\varepsilon'\right)}.$$

Of course, the product of supermodular functions is not necessary supermodular. Notice that

$$\begin{aligned}
&\frac{\partial^2\left[S_t\left(k,k',\varepsilon,\varepsilon'\right)m\left(\varepsilon,\varepsilon'\right)\right]}{\partial\varepsilon\partial\varepsilon'}\\
=\;& m\left(\varepsilon,\varepsilon'\right)\frac{\partial^2 S_t}{\partial\varepsilon\partial\varepsilon'}+\frac{\partial S_t}{\partial\varepsilon}m_2\left(\varepsilon,\varepsilon'\right)+\frac{\partial S_t}{\partial\varepsilon'}m_1\left(\varepsilon,\varepsilon'\right)+S_t m_{12}\left(\varepsilon,\varepsilon'\right)\\
=\;& m\left(\varepsilon,\varepsilon'\right)\frac{\partial^2 S_t}{\partial\varepsilon\partial\varepsilon'}-\kappa'\left(\varepsilon\right)\tilde{\chi}\left(q\right)m_2\left(\varepsilon,\varepsilon'\right)-\kappa'\left(\varepsilon'\right)\tilde{\chi}\left(q\right)m_1\left(\varepsilon,\varepsilon'\right)+S_t m_{12}\left(\varepsilon,\varepsilon'\right)\\
=\;& m\left(\varepsilon,\varepsilon'\right)\frac{\partial^2 S_t}{\partial\varepsilon\partial\varepsilon'}-\frac{\theta_\kappa\left(\varepsilon\right)\kappa\left(\varepsilon\right)}{\varepsilon}\tilde{\chi}\left(q\right)\frac{m_{12}\left(\varepsilon,\varepsilon'\right)\varepsilon}{\theta_m\left(\varepsilon\right)}-\frac{\theta_\kappa\left(\varepsilon'\right)\kappa\left(\varepsilon'\right)}{\varepsilon'}\tilde{\chi}\left(q\right)\frac{m_{12}\left(\varepsilon,\varepsilon'\right)\varepsilon'}{\theta_m\left(\varepsilon'\right)}\\
+\;& S_t m_{12}\left(\varepsilon,\varepsilon'\right)\\
\geq\;& m\left(\varepsilon,\varepsilon'\right)\frac{\partial^2 S_t}{\partial\varepsilon\partial\varepsilon'}+\left[S_t-\kappa\left(\varepsilon\right)\chi\left(q\right)-\kappa\left(\varepsilon'\right)\tilde{\chi}\left(q\right)\right]m_{12}\left(\varepsilon,\varepsilon'\right)\\
\geq\;& 0
\end{aligned}$$

where the last second inequality applies $\theta_\kappa\left(\varepsilon\right)\leq\theta_m\left(\varepsilon\right)$, and the last inequality applies $\frac{\partial^2 S_t}{\partial\varepsilon\partial\varepsilon'}\geq 0$ and $S_t-\kappa\left(\varepsilon\right)\chi\left(q\right)-\kappa\left(\varepsilon'\right)\tilde{\chi}\left(q\right)\geq 0$. **Q.E.D.**

2.C.5 Derivation of Equation (2.14)

Following Üslü (2019), the planner's current-value Hamiltonian can be written as

$$\begin{aligned}\mathcal{H}_t^p \equiv & \int u(k)\, dF_t^p(k) - \int\int \chi\left[\varepsilon_t^p(k), q_t^p(k,k')\right] m\left[\varepsilon_t^p(k), \varepsilon_t^p(k')\right] dF_t^p(k')\, dF_t^p(k) \\ & + \int\int m\left[\varepsilon_t^p(k), \varepsilon_t^p(k')\right] \left\{V_t^p\left[k + q_t^p(k,k')\right] - V_t^p(k)\right\} dF_t^p(k')\, dF_t^p(k) \\ & + \int\int \eta_t(k,k') \left[q_t^p(k,k') + q_t^p(k',k)\right] dF_t^p(k')\, dF_t^p(k). \end{aligned} \tag{2.56}$$

First-order conditions. First, take any optimal q_t^e and

$$\begin{aligned}\hat{q}_t(k,k') &= q_t^e(k,k') + \alpha_q \mathbf{1}\left\{V_t^e(k) > V_t^e(k')\right\} - \alpha_q \mathbf{1}\left\{V_t^e(k) < V_t^e(k')\right\} \\ &= q_t^e(k,k') + \alpha_q \Delta_t(k,k'),\end{aligned}$$

where α_q is an arbitrary scalar. Second, take any optimal $\varepsilon_t^e(k)$, an arbitrary admissible deviation $\delta_t(k)$ and a scalar α_ε, let $\hat{\varepsilon}_t(k) = \varepsilon_t^e(k) + \alpha_\varepsilon \cdot \delta_t(k)$. For small α_q and α_ε, we obtain up to second-order terms:

$$\begin{aligned}& \mathcal{H}_t^p(\hat{\varepsilon}_t, \hat{q}_t) - \mathcal{H}_t^p(\varepsilon_t^e, q_t^e) \\ = \; & -\alpha_\varepsilon \int\int \left\{ \begin{array}{l} \chi_1\left[\varepsilon_t^e(k), q_t^e(k,k')\right] m\left[\varepsilon_t^e(k), \varepsilon_t^e(k')\right] \delta_t(k) \\ +\chi\left[\varepsilon_t^e(k), q_t^e(k,k')\right] m_1\left[\varepsilon_t^e(k), \varepsilon_t^e(k')\right] \delta_t(k) \\ +\chi\left[\varepsilon_t^e(k), q_t^e(k,k')\right] m_2\left[\varepsilon_t^e(k), \varepsilon_t^e(k')\right] \delta_t(k') \end{array} \right\} dF_t^p(k')\, dF_t^p(k) \\ & +\alpha_\varepsilon \int\int \left\{ \begin{array}{l} m_1\left[\varepsilon_t^e(k), \varepsilon_t^e(k')\right] \delta_t(k) \\ +m_2\left[\varepsilon_t^e(k), \varepsilon_t^e(k')\right] \delta_t(k') \end{array} \right\} \left\{V_t^p\left[k + q_t^e(k,k')\right] - V_t^p(k)\right\} dF_t^p(k')\, dF_t^p(k) \\ & -\alpha_q \int\int \chi_2\left[\varepsilon_t^e(k), q_t^e(k,k')\right] m\left[\varepsilon_t^e(k), \varepsilon_t^e(k')\right] \Delta_t(k,k')\, dF_t^p(k')\, dF_t^p(k) \\ & +\alpha_q \int\int m\left[\varepsilon_t^e(k), \varepsilon_t^e(k')\right] V_t^{p\prime}\left[k + q_t^e(k,k')\right] \Delta_t(k,k')\, dF_t^p(k')\, dF_t^p(k) \\ & +\alpha_q \int\int \eta_t(k,k') \left[\Delta_t(k,k') + \Delta_t(k',k)\right] dF_t^p(k')\, dF_t^p(k).\end{aligned}$$

We can rewrite the above equation as

$$
\begin{aligned}
& \mathcal{H}_t^p\left(\hat{\varepsilon}_t, \hat{q}_t\right)-\mathcal{H}_t^p\left(\varepsilon_t^e, q_t^e\right) \\
= & \alpha_\varepsilon \int\int \left\{ \begin{array}{c} m_1\left[\varepsilon_t^e(k), \varepsilon_t^e\left(k'\right)\right]\left\{V_t^p\left[k+q_t^e\left(k, k'\right)\right]-V_t^p(k)\right\} \\ +m_2\left[\varepsilon_t^e\left(k'\right), \varepsilon_t^e(k)\right]\left\{V_t^p\left[k'-q_t^e\left(k, k'\right)\right]-V_t^p\left(k'\right)\right\} \\ -\chi_1\left[\varepsilon_t^e(k), q_t^e\left(k, k'\right)\right] m\left[\varepsilon_t^e(k), \varepsilon_t^e\left(k'\right)\right] \\ -\chi\left[\varepsilon_t^e(k), q_t^e\left(k, k'\right)\right] m_1\left[\varepsilon_t^e(k), \varepsilon_t^e\left(k'\right)\right] \\ -\chi\left[\varepsilon_t^e\left(k'\right),-q_t^e\left(k, k'\right)\right] m_2\left[\varepsilon_t^e\left(k'\right), \varepsilon_t^e(k)\right] \end{array} \right\} \delta_t(k) dF_t^p\left(k'\right) dF_t^p(k) \\
& +\frac{\alpha_q}{2} \int\int m\left[\varepsilon_t^e(k), \varepsilon_t^e\left(k'\right)\right]\left\{V_t^{p\prime}\left[k+q_t^e\left(k, k'\right)\right]-\chi_2\left[\varepsilon_t^e(k), q_t^e\left(k, k'\right)\right]\right\} \Delta_t\left(k, k'\right) \\
& \times dF_t^p\left(k'\right) dF_t^p(k) \\
& +\frac{\alpha_q}{2} \int\int m\left[\varepsilon_t^e(k), \varepsilon_t^e\left(k'\right)\right]\left\{V_t^{p\prime}\left[k'+q_t^e\left(k', k\right)\right]-\chi_2\left[\varepsilon_t^e\left(k'\right), q_t^e\left(k', k\right)\right]\right\} \Delta_t\left(k', k\right) \\
& \times dF_t^p\left(k'\right) dF_t^p(k) \\
= & \alpha_\varepsilon \int\int \left\{ \begin{array}{c} m_1\left[\varepsilon_t^e(k), \varepsilon_t^e\left(k'\right)\right]\left\{V_t^p\left[k+q_t^p\left(k, k'\right)\right]-V_t^p(k)\right\} \\ +m_1\left[\varepsilon_t^e(k), \varepsilon_t^e\left(k'\right)\right]\left\{V_t^p\left[k'-q_t^p\left(k, k'\right)\right]-V_t^p\left(k'\right)\right\} \\ -\chi_1\left[\varepsilon_t^e(k), q_t^p\left(k, k'\right)\right] m\left[\varepsilon_t^e(k), \varepsilon_t^e\left(k'\right)\right] \\ -\chi\left[\varepsilon_t^e(k), q_t^p\left(k, k'\right)\right] m_1\left[\varepsilon_t^e(k), \varepsilon_t^e\left(k'\right)\right] \\ -\chi\left[\varepsilon_t^e\left(k'\right),-q_t^p\left(k, k'\right)\right] m_1\left[\varepsilon_t^e(k), \varepsilon_t^e\left(k'\right)\right] \end{array} \right\} \delta_t(k) dF_t^p\left(k'\right) dF_t^p(k) \\
& +\frac{\alpha_q}{2} \int\int m\left[\varepsilon_t^e(k), \varepsilon_t^e\left(k'\right)\right] \left\{ \begin{array}{c} V_t^{p\prime}\left[k+q_t^e\left(k, k'\right)\right]-V_t^{p'}\left[k'-q_t^e\left(k, k'\right)\right] \\ -\chi_2\left[\varepsilon_t^e(k), q_t^e\left(k, k'\right)\right]+\chi_2\left[\varepsilon_t^e\left(k'\right),-q_t^e\left(k, k'\right)\right] \end{array} \right\} \\
& \times \Delta_t\left(k, k'\right) dF_t^p\left(k'\right) dF_t^p(k),
\end{aligned}
$$

where we apply $\Delta_t\left(k, k'\right)+\Delta_t\left(k', k\right)=0$ in the first and second equality and $q_t^e\left(k, k'\right)+q_t^e\left(k', k\right)=0$ in the second equality.

If $\left\{\varepsilon_t^e, q_t^e\right\}$ is optimal, the above equation must be negative. Thus the integrand in the second term must be zero everywhere. Then the FOC for $q_t^e\left(k, k'\right)$ becomes

$$
V_t^{p\prime}\left[k+q_t^e\left(k, k'\right)\right]-V_t^{p'}\left[k'-q_t^e\left(k, k'\right)\right]-\chi_2\left[\varepsilon_t^e(k), q_t^e\left(k, k'\right)\right]+\chi_2\left[\varepsilon_t^e\left(k'\right),-q_t^e\left(k, k'\right)\right]=0.
$$

In other words, $q_t^p(k,k')$ is the solution to

$$q_t^p(k,k') = \arg\max_q \left\{ V_t^p(k+q) + V_t^p(k'-q) - \chi\left(\varepsilon_t^e(k), q\right) - \chi\left(\varepsilon_t^e(k'), -q\right) \right\}.$$

Moreover, for the FOC of ε_t^e, since $\delta_t(k)$ is an arbitrary admissible deviation, we must have

$$\begin{aligned}
& m_1\left[\varepsilon_t^e(k), \varepsilon_t^e(k')\right] \left\{ \begin{array}{c} V_t^p\left[k + q_t^p(k,k')\right] - V_t^p(k) - \chi\left[\varepsilon_t^e(k), q_t^p(k,k')\right] \\ +V_t^p\left[k' - q_t^p(k,k')\right] - V_t^p(k') - \chi\left[\varepsilon_t^e(k'), -q_t^p(k,k')\right] \end{array} \right\} \\
& -\chi_1\left[\varepsilon_t^e(k), q_t^p(k,k')\right] m\left[\varepsilon_t^e(k), \varepsilon_t^e(k')\right] \\
& \left\{ \begin{array}{ll} \leq 0, & \text{if } \varepsilon_t^e(k) = 0, \\ = 0, & \text{if } \varepsilon_t^e(k) \in (0,1), \quad , \\ \geq 0, & \text{if } \varepsilon_t^e(k) = 1. \end{array} \right.
\end{aligned}$$

Thus the constrained efficiency solution of ε_t^p must satisfy

$$\Gamma_t^p\left(\varepsilon_t^p\right)(k) \equiv \arg\max_{\varepsilon \in [0,1]} \left\{ \int S_t^p\left(k, k', \varepsilon, \varepsilon_t^p(k')\right) m\left[\varepsilon, \varepsilon_t^p(k')\right] dF_t(k') \right\},$$

where

$$\begin{aligned}
S_t^p(k,k',\varepsilon,\varepsilon') &= V_t^p\left[k + q_t^p(k,k')\right] - V_t^p(k) - \chi\left[\varepsilon, q_t^p(k,k')\right] \\
&\quad + V_t^p\left[k' - q_t^p(k,k')\right] - V_t^p(k') - \chi\left[\varepsilon', -q_t^p(k,k')\right].
\end{aligned}$$

2.C.6 Proof of Proposition 2.3

Proof. Denote v_t^w as the co-state to a_t, the Hamiltonian is thus given by

$$\mathcal{H}_t^w \equiv u\left(\frac{a_t}{1+\rho_t^w}\right) - e^{-r(T+\Delta-t)} d\delta_t + v_t^w\left(\frac{\dot{\rho}_t^w}{1+\rho_t^w} a_t + d\delta_t\right). \tag{2.57}$$

The evolution of costate is given by $rv_t^w - \dot{v}_t^w = \frac{\partial \mathcal{H}_t^w}{\partial a_t}$, i.e.

$$\dot{v}_t^w = rv_t^w - \frac{1}{1+\rho_t^w} u'\left(\frac{a_t}{1+\rho_t^w}\right) - v_t^w \frac{\dot{\rho}_t^w}{1+\rho_t^w}. \tag{2.58}$$

The first order condition with respect to $d\delta_t$ is

$$v_t^w = e^{-r(T+\Delta-t)}. \tag{2.59}$$

Since the first order condition is independent to a_t and δ_t, all banks must have the same value of costate. But since the evolution of costate, 2.58, depends on a_t, the only possibility is that all banks have the same a_t for all $t > 0$. This implies $\delta_t(a)$ is given by result (b), such that they hold K units of reserve balance for all $t > 0$. Substituting (2.59) to the evolution of costate, (2.58), we have

$$\dot{\rho}_t^w = -e^{r(T+\Delta-t)} u'(K).$$

The solution to the above ODE is

$$\rho_t^w = \rho_T^w + e^{r\Delta}\left[e^{r(T-t)} - 1\right] \frac{u'(K)}{r}.$$

Notice that at T the bank problem is

$$\max_{q_T} \left\{U(k+q_T) - e^{-r\Delta}(1+\rho_T^w) q_T\right\}.$$

To yield $k + q_T = K$, we have

$$\rho_T^w = e^{r\Delta} U'(K) - 1.$$

Q.E.D.

2.C.7 Proof of Proposition 2.4

Proof. [$\varepsilon_t(k) = 0$ is always an equilibrium] For any k and ε, if all the other banks choose zero search intensity, then

$$\frac{\partial \int S_t(k,k',\varepsilon,0)\, m(\varepsilon,0)\, dF_t(k')}{\partial \varepsilon} = \int \frac{-\kappa\lambda_0 \left[(k'-k)\, V_t''(k)\right]^2}{4\left[\kappa(\varepsilon) - \frac{1}{2}\left(V_t''(k) + V_t''(k')\right)\right]^2} dF_t(k') < 0.$$

This implies that the bank k's optimal response is $\varepsilon = 0$. Thus $\varepsilon_t(k) = 0\ \forall k$ is a self-fulfilling equilibrium.

[Possibility of multiple equilibria] To show that it is possible to have multiple equilibria under some parameter conditions, we provide a necessary and sufficient condition for $\varepsilon_t(k) = 1\ \forall k$ to be an equilibrium. Suppose all the other banks choose search intensity $\varepsilon_t = 1$. Then for any k and ε, we have

$$\begin{aligned}
&\frac{\partial \int S_t(k,k',\varepsilon,1)\, m(\varepsilon,1)\, dF_t(k')}{\partial \varepsilon} \\
= &\int \left[\frac{k'-k}{2} V_t''(k)\right]^2 \frac{(\lambda-\lambda_0)\kappa - \frac{\lambda-\lambda_0}{2}\left[V_t''(k) + V_t''(k')\right] - \kappa\lambda_0}{\left[\kappa(\varepsilon+1) - \frac{1}{2}\left(V_t''(k) + V_t''(k')\right)\right]^2} dF_t(k') \\
= &\frac{(\lambda-\lambda_0)\kappa - (\lambda-\lambda_0) V_t''(k) - \kappa\lambda_0}{\left[\kappa(\varepsilon+1) - V_t''(k)\right]^2} \left(\frac{V_t''(k)}{2}\right)^2 \int (k'-k)^2\, dF_t(k'),
\end{aligned}$$

where the second equality is because $V_t''(k)$ is a constant in k. Thus the sufficient and necessary condition for $\varepsilon_t(k) = 1\ \forall k$ to be an equilibrium is that $V_t''(k) \leq \frac{\lambda-2\lambda_0}{\lambda-\lambda_0}\kappa$.

[The largest equilibrium is either $\varepsilon_t(k) = 0\ \forall k$ or $\varepsilon_t(k) = 1\ \forall k$] Let $\varepsilon_t^{\max}(k)$ be the largest equilibrium search profile. Denote $\bar{\varepsilon}_t = \sup_k \{\varepsilon_t^{\max}(k)\}$ and $\underline{\varepsilon}_t = \inf_k \{\varepsilon_t^{\max}(k)\}$. We first prove $\bar{\varepsilon}_t = \underline{\varepsilon}_t$ by contradiction. Suppose $\bar{\varepsilon}_t > \underline{\varepsilon}_t$, then $\frac{\partial^2 S_t(k,k',\varepsilon,\varepsilon') m(\varepsilon,\varepsilon')}{\partial\varepsilon\partial\varepsilon'} > 0$ implies

$$\begin{aligned}
&\frac{\partial \int S_t(k,k',\varepsilon_t^{\max}(k), \bar{\varepsilon}_t)\, m(\varepsilon_t^{\max}(k), \bar{\varepsilon}_t)\, dF_t(k')}{\partial \varepsilon} \\
\geq\ &\frac{\partial \int S_t(k,k',\varepsilon_t^{\max}(k), \varepsilon_t^{\max}(k'))\, m(\varepsilon_t^{\max}(k), \varepsilon_t^{\max}(k'))\, dF_t(k')}{\partial \varepsilon} \geq 0
\end{aligned} \tag{2.60}$$

for any k such that $\varepsilon_t^{\max}(k) > 0$. Note that for any k and ε,

$$\begin{aligned}
&\frac{\partial \int S_t(k,k',\varepsilon,\bar{\varepsilon}_t)\, m(\varepsilon,\bar{\varepsilon}_t)\, dF_t(k')}{\partial \varepsilon} \\
= &\int \left[\frac{k'-k}{2} V_t''(k)\right]^2 \frac{(\lambda-\lambda_0)\kappa(\bar{\varepsilon}_t)^2 - \frac{\lambda-\lambda_0}{2}\bar{\varepsilon}_t\left[V_t''(k)+V_t''(k')\right] - \kappa\lambda_0}{\left[\kappa(\varepsilon+\bar{\varepsilon}_t) - \frac{1}{2}\left(V_t''(k)+V_t''(k')\right)\right]^2} dF_t(k') \\
= &\frac{(\lambda-\lambda_0)\kappa(\bar{\varepsilon}_t)^2 - (\lambda-\lambda_0)\bar{\varepsilon}_t V_t''(k) - \kappa\lambda_0}{\left[\kappa(\varepsilon+\bar{\varepsilon}_t) - V_t''(k)\right]^2} \left(\frac{V_t''(k)}{2}\right)^2 \int (k'-k)^2 dF_t(k').
\end{aligned}$$

Since $V_t''(k)$ is negative and constant over k, and $\bar{\varepsilon}_t \in [0,1]$, then equation (2.60) implies that $0 \leq (\lambda-\lambda_0)\kappa(\bar{\varepsilon}_t)^2 - (\lambda-\lambda_0)\bar{\varepsilon}_t V_t''(k) - \kappa\lambda_0 \leq (\lambda-\lambda_0)\kappa - (\lambda-\lambda_0)V_t''(k) - \kappa\lambda_0$ for any k. Then we have

$$\frac{\partial \int S_t(k,k',\varepsilon,1)\, m(\varepsilon,1)\, dF_t(k')}{\partial \varepsilon} \geq 0 \text{ for any } k \text{ and } \varepsilon.$$

Thus there exists an equilibrium search profile where $\varepsilon_t(k) \equiv 1$. Apparently this search profile dominates $\varepsilon_t^{\max}(k)$, which is a contradiction.

Next we prove $\bar{\varepsilon}_t = 0$ or 1 by contradiction. Suppose not, i.e. $\bar{\varepsilon}_t = \underline{\varepsilon}_t = \hat{\varepsilon} \in (0,1)$. Then $\frac{\partial^2 S_t(k,k',\varepsilon,\varepsilon')m(\varepsilon,\varepsilon')}{\partial\varepsilon\partial\varepsilon'} > 0$ implies that for any k, ε,

$$\begin{aligned}
&\frac{\partial \int S_t(k,k',\varepsilon,1)\, m(\varepsilon,1)\, dF_t(k')}{\partial \varepsilon} \\
> &\frac{\partial \int S_t(k,k',\varepsilon,\hat{\varepsilon})\, m(\varepsilon,\hat{\varepsilon})\, dF_t(k')}{\partial \varepsilon} \propto \frac{\partial \int S_t(k,k',\hat{\varepsilon},\hat{\varepsilon})\, m(\varepsilon,\hat{\varepsilon})\, dF_t(k')}{\partial \varepsilon} = 0.
\end{aligned}$$

Thus there exists an equilibrium search profile where $\varepsilon_t(k) \equiv 1$, which is a contradiction. **Q.E.D.**

2.C.8 Proof of Proposition 2.5

Proof. Given $\{F_t\}$, the value function satisfying (2.8) is unique. Guess that

$$V_t(k) = -H_t k^2 + E_t k + D_t. \tag{2.61}$$

Then we have

$$V_t(k+q) - V_t(k) - \chi(\varepsilon, q) = [E_t - 2H_t k - (H_t + \kappa\varepsilon) q] q.$$

The bargaining solution thus solves

$$\begin{aligned} q_t(k, k', \varepsilon, \varepsilon') &= \arg\max_q \{V_t(k+q) + V_t(k'-q) - \chi(\varepsilon, q) - \chi(\varepsilon', q)\}, \\ &= \arg\max_q \left\{-H_t(k+q)^2 - H_t(k'-q)^2 - \kappa(\varepsilon+\varepsilon') q^2\right\}, \\ &= \frac{H_t(k'-k)}{\kappa(\varepsilon+\varepsilon') + 2H_t}, \end{aligned}$$

and

$$\begin{aligned} e^{-r(T+\Delta-t)} R_t(k, k', \varepsilon, \varepsilon') &= \frac{1}{2}\begin{bmatrix} V_t[k + q_t(k, k', \varepsilon, \varepsilon')] - V_t(k) - \chi[\varepsilon, q_t(k, k', \varepsilon, \varepsilon')] \\ V_t(k') - V_t[k' - q_t(k, k', \varepsilon, \varepsilon')] + \chi[\varepsilon', -q_t(k, k', \varepsilon, \varepsilon')] \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} E_t - 2H_t k - (H_t + \kappa\varepsilon) q_t(k, k', \varepsilon, \varepsilon') \\ +E_t - 2H_t k' + (H_t + \kappa\varepsilon') q_t(k, k', \varepsilon, \varepsilon') \end{bmatrix} q_t(k, k', \varepsilon, \varepsilon') \\ &= \left[E_t - H_t(k+k') - \frac{\kappa(\varepsilon - \varepsilon')}{2} q_t(k, k', \varepsilon, \varepsilon')\right] q_t(k, k', \varepsilon, \varepsilon'). \end{aligned}$$

Thus the bilateral Federal funds rate is

$$1 + \rho_t(k, k', \varepsilon, \varepsilon') = \frac{R_t(k, k', \varepsilon, \varepsilon')}{q_t(k, k', \varepsilon, \varepsilon')} = e^{r(T+\Delta-t)}\left[E_t - H_t(k+k') - \frac{\kappa(\varepsilon - \varepsilon')}{2} q_t(k, k', \varepsilon, \varepsilon')\right].$$

The trade surplus is given by

$$\begin{aligned} S_t(k,k',\varepsilon,\varepsilon') &\equiv V_t\left[k+q_t(k,k',\varepsilon,\varepsilon')\right]-V_t(k)-\chi\left[\varepsilon,q_t(k,k',\varepsilon,\varepsilon')\right] \\ &\quad +V_t\left[k'-q_t(k,k',\varepsilon,\varepsilon')\right]-V_t(k')-\chi\left[\varepsilon',-q_t(k,k',\varepsilon,\varepsilon')\right], \\ &= -H_t\left\{\left[k+\frac{H_t(k'-k)}{\kappa(\varepsilon+\varepsilon')+2H_t}\right]^2+\left[k'-\frac{H_t(k'-k)}{\kappa(\varepsilon+\varepsilon')+2H_t}\right]^2-k^2-k'^2\right\} \\ &\quad -\kappa(\varepsilon+\varepsilon')\left[\frac{H_t(k'-k)}{\kappa(\varepsilon+\varepsilon')+2H_t}\right]^2, \\ &= \frac{\left[H_t(k'-k)\right]^2}{\kappa(\varepsilon+\varepsilon')+2H_t}. \end{aligned}$$

The equilibrium search profile is a fixed point function to the following functional:

$$\begin{aligned} \Gamma_t(\varepsilon_t)(k) &\equiv \arg\max_{\varepsilon\in[0,1]}\left\{\int S_t\left[k,k',\varepsilon,\varepsilon_t(k')\right]m\left[\varepsilon,\varepsilon_t(k')\right]dF_t(k')\right\} \\ &= \arg\max_{\varepsilon\in[0,1]}\left\{\int\frac{\left[H_t(k'-k)\right]^2}{\kappa\left[\varepsilon+\varepsilon_t(k')\right]+2H_t}\left[(\lambda-\lambda_0)\varepsilon\varepsilon_t(k')+\lambda_0\right]dF_t(k')\right\} \\ &= \arg\max_{\varepsilon\in[0,1]}\left\{\frac{(H_t)^2}{\kappa(\varepsilon+\varepsilon_t)+2H_t}\left[(\lambda-\lambda_0)\varepsilon\varepsilon_t+\lambda_0\right]\int(k'-k)^2dF_t(k')\right\} \end{aligned} \tag{2.62}$$

which only depends on H_t and F_t and, is independent of k and k'. The last equality is guaranteed by Proposition 2.4. Thus, we write $\Gamma_t(\varepsilon)(k)=\Gamma(\varepsilon;H_t)$, where the latter is given by (2.20). The equilibrium search intensity at t is the fixed point of $\Gamma(\varepsilon_t;H_t)$, which is any element of $\Omega(h)$. The HJB equation becomes

$$\begin{aligned} r\left[-H_tk^2+E_tk+D_t\right] &= -\dot{H}_tk^2+\dot{E}_tk+\dot{D}_t-a_2k^2+a_1k \\ &\quad +\frac{1}{4}\frac{H_t^2}{\kappa\varepsilon_t+H_t}\left[(\lambda-\lambda_0)\varepsilon_t^2+\lambda_0\right]\int(k'-k)^2dF_t(k'). \end{aligned}$$

Matching the coefficients, we have

$$\begin{aligned} rH_t &= \dot{H}_t + a_2 - \frac{1}{4}\frac{H_t^2}{\kappa\varepsilon_t + H_t}\left[(\lambda-\lambda_0)\,\varepsilon_t^2 + \lambda_0\right], \\ rE_t &= \dot{E}_t + a_1 - \frac{1}{2}\frac{H_t^2 K}{\kappa\varepsilon_t + H_t}\left[(\lambda-\lambda_0)\,\varepsilon_t^2 + \lambda_0\right], \\ rD_t &= \dot{D}_t + \frac{1}{4}\frac{H_t^2}{\kappa\varepsilon_t + H_t}\left[(\lambda-\lambda_0)\,\varepsilon_t^2 + \lambda_0\right]\int k'^2 dF_t\left(k'\right), \end{aligned}$$

where the fact that $V_T(k) = -A_2k^2 + A_1k$ implies the terminal conditions

$$H_T = A_2,\ E_T = A_1,\ D_T = 0.$$

Q.E.D.

2.C.9 Proof of Proposition 2.6

Proof. Notice that the first-order condition of bank k's search intensity is

$$\begin{aligned} &\frac{\partial \int S_t\left(k,k',\varepsilon,\varepsilon_t\right) m\left(\varepsilon,\varepsilon_t\right) dF_t\left(k'\right)}{\partial\varepsilon} \\ =\ &\frac{(\lambda-\lambda_0)\,\kappa\left(\varepsilon_t\right)^2 + 2\left(\lambda-\lambda_0\right)\varepsilon_t H_t - \kappa\lambda_0}{\left[\kappa\left(\varepsilon+\bar{\varepsilon}_t\right)+2H_t\right]^2}\left(H_t\right)^2\int\left(k'-k\right)^2 dF_t\left(k'\right). \end{aligned}$$

Thus $\varepsilon = 1$ if $(\lambda-\lambda_0)\,\kappa\left(\varepsilon_t\right)^2 + 2\left(\lambda-\lambda_0\right)\varepsilon_t H_t - \kappa\lambda_0 > 0$. It implies that the largest equilibrium satisfies

$$\Gamma_t\left(\varepsilon_t\right)(k)\begin{cases} = 1,\ \text{if } H_t > \frac{[\lambda_0-(\lambda-\lambda_0)]\kappa}{2(\lambda-\lambda_0)}; \\ 0,\ \text{otherwise.} \end{cases}$$

By rearranging the terms, we obtain the proposition. **Q.E.D.**

2.C.10 Proof of Lemma 2.2

Proof. We first derive the expressions of μ_1, μ_2, $\tau_1(H; A, u)$, $J(t; A, u)$, and $\tau_2(H; A, u)$ by solving the ODEs of H_t with a terminal value $H_u = A$ under $\varepsilon_t = 1$ and $\varepsilon_t = 0$. Then we characterize the equilibrium dynamics.

[Solve the ODE of H_t under $\varepsilon_t = 1$] If $\varepsilon_t = 1$, the law of motion of H_t is

$$\begin{aligned} \dot{H}_t &= rH_t - a_2 + \frac{\lambda}{4}\frac{H_t^2}{\kappa + H_t} = \frac{(4r+\lambda)H_t^2 + 4(\kappa r - a_2)H_t - 4\kappa a_2}{4(\kappa + H_t)} \\ &= \frac{4r+\lambda}{4(\kappa + H_t)}(H_t - \mu_1)(H_t - \mu_2), \end{aligned} \tag{2.63}$$

where μ_1 and μ_2 are the zero point of formula $(4r+\lambda)H_t^2 + 4(\kappa r - a_2)H_t - 4\kappa a_2 = 0$, and they are given by

$$\begin{aligned} \mu_1 &\equiv \frac{1}{2r + \frac{\lambda}{2}}\left\{-(\kappa r - a_2) - \left[(\kappa r - a_2)^2 + a_2\kappa(4r+\lambda)\right]^{0.5}\right\}, \\ \mu_2 &\equiv \frac{1}{2r + \frac{\lambda}{2}}\left\{-(\kappa r - a_2) + \left[(\kappa r - a_2)^2 + a_2\kappa(4r+\lambda)\right]^{0.5}\right\}. \end{aligned}$$

Equation (2.63) can be written as

$$\begin{aligned} \frac{4r+\lambda}{4}dt &= \frac{\kappa + H_t}{(H_t - \mu_1)(H_t - \mu_2)}dH_t \\ &= \frac{\kappa + \mu_1}{\mu_1 - \mu_2}\cdot\frac{1}{H_t - \mu_1}dH_t - \frac{\kappa + \mu_2}{\mu_1 - \mu_2}\cdot\frac{1}{H_t - \mu_2}dH_t \end{aligned}$$

Given a terminal value condition $H_u = A$, then H_t satisfies

$$\left(r + \frac{\lambda}{4}\right)(u - t) = \frac{\kappa + \mu_1}{\mu_1 - \mu_2}\cdot\log\left(\frac{A - \mu_1}{H_t - \mu_1}\right) - \frac{\kappa + \mu_2}{\mu_1 - \mu_2}\cdot\log\left(\frac{A - \mu_2}{H_t - \mu_2}\right).$$

By rearraning the terms, we can write t as a function of H_t:

$$t = \tau_1 (H_t; A, u) = u - \frac{(\kappa + \mu_1) \log \left(\frac{A - \mu_1}{H_t - \mu_1} \right) - (\kappa + \mu_2) \log \left(\frac{A - \mu_2}{H_t - \mu_2} \right)}{\left(r + \frac{\lambda}{4} \right) (\mu_1 - \mu_2)}.$$

[Solve the ODE of H_t under $\varepsilon_t = 0$] If $\varepsilon_t = 0$, the law of motion of H_t is

$$\dot{H}_t = r H_t - a_2 + \frac{\lambda_0}{4} H_t = \left(r + \frac{\lambda_0}{4} \right) H_t - a_2. \tag{2.64}$$

Given a terminal value condition $H_u = A$, the solution to H_t is

$$H_t = J (t; A, u) = \frac{a_2}{r + \frac{\lambda_0}{4}} + \left(A - \frac{a_2}{r + \frac{\lambda_0}{4}} \right) e^{-\left(r + \frac{\lambda_0}{4} \right)(u - t)}.$$

Note that $J (t; A, u)$ is monotone in t. Thus we can get the inverse function:

$$t = \tau_2 (H; A, u) = u - \frac{1}{r + \frac{\lambda_0}{4}} \log \left(\frac{A - \frac{a_2}{r + \frac{\lambda_0}{4}}}{H - \frac{a_2}{r + \frac{\lambda_0}{4}}} \right) = u + \frac{1}{r + \frac{\lambda_0}{4}} \log \left(1 - \frac{H - A}{\frac{a_2}{r + \frac{\lambda_0}{4}} - A} \right).$$

[Characterize the dynamics of ε_t and H_t] Since the terminal value of H_T is fixed, we characterize the time paths of ε_t and H_t inversely from T to 0. The characterization is divided into the following cases.

Case 1.1: $A_2 \geq \eta$, $\dot{H}_t \Big|_{H_t = \eta, \varepsilon_t = 1} > 0$, $\dot{H}_t \Big|_{H_t = \eta, \varepsilon_t = 0} > 0$ **and** $\tau_1 (H_t; A, u) > 0$**.** In this case, as t decreases from T, the equilibrium solution of ε_t and H_t is given by $\varepsilon_t = 1$ and $t = \tau_1 (H_t; A_2, T)$. Moreover, according to equation (2.63), $\dot{H}_t \Big|_{H_t = \eta, \varepsilon_t = 1} > 0$ guarantees that H_t is decreasing as t goes from T to 0 before hitting η. According to the definition of $\tau_1 (H_t; A, u)$, the time of H_t hitting η is $\tau_1 (\eta; A_2, T)$. A positive $\tau_1 (\eta; A_2, T)$ means that H_t decreases to η before time 0. The condition $\dot{H}_t \Big|_{H_t = \eta, \varepsilon_t = 0} > 0$ guarantees that after H_t hits η, H_t continues to decrease as t goes to 0 and $\varepsilon_t = 0$ for $t < \tau_1 (\eta; A_2, T)$. Note that

the necessary and sufficient parameter conditions for $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=1} > 0$ and $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=0} > 0$ are $a_2 < r\eta + \frac{\lambda}{4}\frac{\eta^2}{\kappa+\eta} = \left(r - \frac{\lambda}{4} + \frac{\lambda_0}{2}\right)\eta$ and $a_2 < \left(r + \frac{\lambda_0}{4}\right)\eta$, respectively. Since $a_2 \geq 0$, then we have $\left(r + \frac{\lambda_0}{4}\right)\eta > \left(r - \frac{\lambda}{4} + \frac{\lambda_0}{2}\right)\eta$. Therefore, when $A_2 \geq \eta$, $a_2 < \left(r - \frac{\lambda}{4} + \frac{\lambda_0}{2}\right)\eta$ and $\tau_1(\eta; A_2, T) > 0$, the paths of ε_t and H_t are

$$\varepsilon_t = \begin{cases} 1, \text{ if } t \geq \tau_1(\eta; A_2, T); \\ 0, \text{ otherwise.} \end{cases}$$

$$H_t = \begin{cases} \tau_1^{-1}(t; A_2, T), \text{ if } t \geq \tau_1(\eta; A_2, T); \\ J[t; \eta, \tau_1(\eta; A_2, T)], \text{ otherwise.} \end{cases}$$

Case 1.2: $A_2 \geq \eta$, $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=1} > 0$, $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=0} \leq 0$ **and** $\tau_1(H_t; A, u) > 0$**.** In this case, when hitting η at $\tau_1(\eta; A_2, T)$, H_t will stay at η until time 0. This is because when $H_t < \eta$, $\dot{H}_t\Big|_{\varepsilon_t=0} < \dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=0} \leq 0$. Therefore, when $A_2 \geq \eta$, $a_2 \in \left[\left(r + \frac{\lambda_0}{4}\right)\eta, \left(r - \frac{\lambda}{4} + \frac{\lambda_0}{2}\right)\eta\right)$ and $\tau_1(\eta; A_2, T) > 0$, the paths of ε_t and H_t are $\varepsilon_t = 1$ for all $t \in [0, T]$ and

$$H_t = \begin{cases} \tau_1^{-1}(t; A_2, T), \text{ if } t \geq \tau_1(\eta; A_2, T); \\ \eta, \text{ otherwise.} \end{cases}$$

However, this case doesn't exist due to the following reason. If $\eta > 0$, then we have $\left(r + \frac{\lambda_0}{4}\right)\eta > \left(r - \frac{\lambda}{4} + \frac{\lambda_0}{2}\right)\eta$, which implies that the condition $a_2 \in \left[\left(r + \frac{\lambda_0}{4}\right)\eta, \left(r - \frac{\lambda}{4} + \frac{\lambda_0}{2}\right)\eta\right)$ is an empty set. If $\eta \leq 0$, then the equilibrium path contradicts with that $H_t > 0$. Combining Case 1.1 and 1.2, we obtain Case (a-i) of the lemma.

Case 1.3: (1) $A_2 \geq \eta$**; (2)** $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=1} \leq 0$ **or** $\tau_1(\eta; A_2, T) \leq 0$**.** This is the counterpart of Case 1.1 and 1.2. If $\tau_1(\eta; A_2, T) \leq 0$, then H_t will never hit η before the time goes to zero. If $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=1} \leq 0$, then $\mu_2 \geq \eta$ and H_t monotonically converges to μ_2 before hitting η. Both conditions imply that $\varepsilon_t = 1$ for all $t \in [0, T]$ and $H_t = \tau_1^{-1}(t; A_2, T)$. This establishes Case (a-ii) in the lemma.

Case 2.1: $A_2 < \eta$, $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=0} < 0$, $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=1} < 0$ **and** $\tau_2(\eta; A_2, T) > 0$**.** In this case, as t decreases from T, the equilibrium solution of ε_t and H_t is $\varepsilon_t = 0$ and $H_t = J(t; A_2, T)$. Moreover, according to equation (2.64), $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=0} < 0$ guarantees that H_t is increasing as t goes from T to 0 before hitting η. According to the definition of $\tau_2(H_t; A, u)$, the time of H_t hitting η is $\tau_2(\eta; A_2, T)$. A positive $\tau_2(\eta; A_2, T)$ means that H_t increases to η before time 0. The condition $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=1} < 0$ guarantees that after H_t hits η, H_t continues to increase as t goes to 0 and $\varepsilon_t = 1$ for $t \leq \tau_2(\eta; A_2, T)$. The necessary and sufficient parameter conditions for $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=0} < 0$ and $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=1} < 0$ are $a_2 > \left(r + \frac{\lambda_0}{4}\right)\eta$ and $a_2 > \left(r - \frac{\lambda}{4} + \frac{\lambda_0}{2}\right)\eta$, respectively. Since $\lambda > \lambda_0$ and $\eta > A_2 > 0$, we have $\left(r + \frac{\lambda_0}{4}\right)\eta > \left(r - \frac{\lambda}{4} + \frac{\lambda_0}{2}\right)\eta$. Therefore, when $A_2 < \eta$, $a_2 > \left(r + \frac{\lambda_0}{4}\right)\eta$ and $\tau_2(\eta; A_2, T) > 0$, the paths of ε_t and H_t are

$$\varepsilon_t = \begin{cases} 0, \text{ if } t > \tau_2(\eta; A_2, T); \\ 1, \text{ otherwise.} \end{cases}$$

$$H_t = \begin{cases} J(t; A_2, T), \text{ if } t \geq \tau_2(\eta; A_2, T); \\ \tau_1^{-1}(t; \eta, \tau_2(\eta; A_2, T)), \text{ otherwise.} \end{cases}$$

Case 2.2: $A_2 < \eta$, $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=0} < 0$, $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=1} \geq 0$ **and** $\tau_2(\eta; A_2, T) > 0$**.** In this case, when hitting η at $\tau_2(\eta; A_2, T)$, H_t will stay at η until time 0. This is because when $H_t > \eta$, $\dot{H}_t\Big|_{\varepsilon_t=1} > \dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=1} \geq 0$. Therefore, when $A_2 < \eta$, $a_2 \in \left(\left(r + \frac{\lambda_0}{4}\right)\eta, \left(r - \frac{\lambda}{4} + \frac{\lambda_0}{2}\right)\eta\right]$ and $\tau_2(\eta; A_2, T) > 0$, the paths of ε_t and H_t are

$$\varepsilon_t = \begin{cases} 0, \text{ if } t > \tau_2(\eta; A_2, T); \\ 1, \text{ otherwise.} \end{cases}$$

$$H_t = \begin{cases} J(t; A_2, T), \text{ if } t \geq \tau_2(\eta; A_2, T); \\ \eta, \text{ otherwise.} \end{cases}$$

However, since $\left(r+\frac{\lambda_0}{4}\right)\eta > \left(r-\frac{\lambda}{4}+\frac{\lambda_0}{2}\right)\eta$, this case doesn't exist. Combining Case 2.1 and 2.2, we obtain Case (b-i) of the lemma.

Case 2.3: (1) $A_2 < \eta$**; (2)** $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=0} \geq 0$ **or** $\tau_2(\eta; A_2, T) \leq 0$**.** This is the counterpart of Case 2.1 and 2.2. If $\tau_2(\eta; A_2, T) \leq 0$, then H_t will never hit η before the time goes to zero. If $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=0} \geq 0$, then $\frac{a_2}{r+\frac{\lambda_0}{4}} \leq \eta$ and H_t monotonically converges to $\frac{a_2}{r+\frac{\lambda_0}{4}}$ before hitting η. Both conditions imply that $\varepsilon_t = 0$ for all $t \in [0,T]$ and $H_t = J(t; A_2, T)$. This establishes Case (b-ii) of the lemma. **Q.E.D.**

2.C.11 Proof of Lemma 2.3

Proof. Plug the closed-form solution (2.24) and $\varepsilon_t(k) \equiv \varepsilon_t$ into the KFE (2.9), we can get

$$\begin{aligned}
\dot{F}_t(k^w) &= m(\varepsilon_t,\varepsilon_t)\left\{\begin{array}{c}\int_{k>k^w}\int 1\left\{k+\frac{H_t(k'-k)}{2\kappa\varepsilon+2H_t}\leq k^w\right\}dF_t(k')\,dF_t(k)\\ -\int_{k\leq k^w}\int 1\left\{k+\frac{H_t(k'-k)}{2\kappa\varepsilon+2H_t}> k^w\right\}dF_t(k')\,dF_t(k)\end{array}\right\}\\
&= m(\varepsilon_t,\varepsilon_t)\left\{\begin{array}{c}\int_{k>k^w} F_t\left[2\left(1+\frac{\kappa\varepsilon_t}{H_t}\right)k^w-\left(1+\frac{2\kappa\varepsilon_t}{H_t}\right)k\right]dF_t(k)\\ -\int_{k\leq k^w}\left[1-F_t\left[2\left(1+\frac{\kappa\varepsilon_t}{H_t}\right)k^w-\left(1+\frac{2\kappa\varepsilon_t}{H_t}\right)k\right]\right]dF_t(k)\end{array}\right\}\\
&= m(\varepsilon_t,\varepsilon_t)\left[\int F_t\left[2\left(1+\frac{\kappa\varepsilon_t}{H_t}\right)k-\left(1+\frac{2\kappa\varepsilon_t}{H_t}\right)k'\right]dF_t(k')-F_t(k)\right].
\end{aligned}$$

Then the probability density function solves the following PDE:

$$\dot{f}_t(k) = m(\varepsilon_t,\varepsilon_t)\left[2\left(1+\frac{\kappa\varepsilon_t}{H_t}\right)\int f_t\left[2\left(1+\frac{\kappa\varepsilon_t}{H_t}\right)k-\left(1+\frac{2\kappa\varepsilon_t}{H_t}\right)k'\right]f_t(k')\,dk'-f_t(k)\right]. \tag{2.65}$$

To characterize the dynamics of moment function, we take advantage of the Fourier transform. We follow the definition of Bracewell (2000) for the Fourier transform:

$$h^*(\nu) = \int e^{-i2\pi\nu x}h(x)\,dx,$$

where $h^*(\cdot)$ is the Fourier transform of the function $h(\cdot)$.

Let $f_t^*(\cdot)$ be the Fourier transform of the equilibrium pdf $f_t(\cdot)$. Then the Fourier transform of equation (2.65) is

$$\dot{f}_t^*(\nu) = m(\varepsilon_t, \varepsilon_t)\left[f_t^*\left(\frac{H_t}{2(H_t+\kappa\varepsilon_t)}\nu\right) f_t^*\left(\frac{H_t+2\kappa\varepsilon_t}{2(H_t+\kappa\varepsilon_t)}\nu\right) - f_t^*(\nu)\right]. \tag{2.66}$$

The PDE (2.66) cannot be solved in closed form. However, it facilitates the calculation fo the moment function which is the derivative of the transform, with respect to ν, at $\nu = 0$. Let us denote $f_t^{*(n)}(\nu)$ be the n-th derivative of $f_t^*(\nu)$ with respect to ν. By taking n-th derivative with respect to ν to both sides of (2.66), we can obtain

$$\begin{aligned} & \dot{f}_t^{*(n)}(\nu) \\ = \; & m(\varepsilon_t, \varepsilon_t)\left[\begin{array}{l} \sum_{i=0}^n C_n^i \frac{(H_t)^{n-i}(H_t+2\kappa\varepsilon_t)^i}{2^n(H_t+\kappa\varepsilon_t)^n} f_t^{*(n-i)}\left(\frac{H_t}{2(H_t+\kappa\varepsilon_t)}\nu\right) f_t^{*(i)}\left(\frac{H_t+2\kappa\varepsilon_t}{2(H_t+\kappa\varepsilon_t)}\nu\right) \\ -f_t^{*(n)}(\nu) \end{array}\right]. \end{aligned} \tag{2.67}$$

Evaluating the above equation at $\nu = 0$, we can get

$$\dot{M}_{n,t} = m(\varepsilon_t, \varepsilon_t)\left[\sum_{i=0}^n C_n^i \frac{(H_t)^{n-i}(H_t+2\kappa\varepsilon_t)^i}{2^n(H_t+\kappa\varepsilon_t)^n} M_{n-i,t}M_{i,t} - M_{n,t}\right].$$

In particular, by definition we have $M_{0,t} = \int f_t(k)\,dk = 1$ and $M_{1,t} = \int k f_t(k)\,dk = K$. Moreover, the second moment of reserve distribution satisfies

$$\dot{M}_{2,t} = m(\varepsilon_t, \varepsilon_t)\left[-\frac{H_t(H_t+2\kappa\varepsilon_t)}{2(H_t+\kappa\varepsilon_t)^2}M_{2,t} + \frac{H_t(H_t+2\kappa\varepsilon_t)}{2(H_t+\kappa\varepsilon_t)^2}K^2\right].$$

Solving this first-order ODE gives rise to the solution (2.33). **Q.E.D.**

2.C.12 Derivations of positive measures of liquidity

Price impact. Note that the terms of trade between k and k' are

$$1+\rho_t(k,k') = e^{r(T+\Delta-t)}\left[E_t - H_t(k+k')\right],$$
$$q_t(k,k') = \frac{H_t(k'-k)}{2(\kappa\varepsilon_t+H_t)} \Rightarrow k' = k + \frac{2(\kappa\varepsilon_t+H_t)}{H_t} q_t(k,k').$$

Therefore, given k and q, we can infer the reserve holding of the counterparty $k'(k,q)$. Thus the Federal funds rate of a bank k that trades reserves q is given by

$$\begin{aligned}
\log(1+\rho_t(k,q)) &= r(T+\Delta-t) + \log\left[E_t - H_t(k+k'(k,q))\right] \\
&= r(T+\Delta-t) + \log\left[E_t - 2kH_t\right] + \log\left[1 - \frac{2(\kappa\varepsilon_t+H_t)}{E_t-2kH_t} q\right] \\
&\approx r(T+\Delta-t) + \log\left[V_t'(k)\right] - \frac{2(\kappa\varepsilon_t+H_t)}{V_t'(k)} q \\
&= r(T+\Delta-t) + \log\left[V_t'(k)\right] - \frac{2(\kappa\varepsilon_t+H_t)}{-2H_t}\frac{q}{k}\frac{kV_t''(k)}{V_t'(k)} \\
&= r(T+\Delta-t) + \log\left[V_t'(k)\right] + \frac{kV_t''(k)}{V_t'(k)} \cdot \frac{q}{k} \cdot \frac{1}{1-\left(1-\frac{\bar{V}''}{2\kappa\varepsilon_t}\right)^{-1}}
\end{aligned}$$

Denote $\theta_{V,t}(k) \equiv -\frac{kV_t''(k)}{V_t'(k)}$ and $\omega_t \equiv \left(1-\frac{\bar{V}''}{2\kappa\varepsilon_t}\right)^{-1}$, we get equation (2.34).

Return reversal. The average Federal funds rate is

$$1+\varrho_t = e^{r(T+\Delta-t)}\left[E_t - 2H_tK\right], \tag{2.68}$$

then the difference between individual Federal funds rate and the average Federal funds rate is

$$\rho_t(k,k') - \varrho_t = e^{r(T+\Delta-t)}(2K-k-k')H_t.$$

Differentiating the rates with respect to time:

$$\dot{\varrho}_t = e^{r(T+\Delta-t)}\left(\dot{E}_t - 2K\dot{H}_t\right) - r\left(1+\varrho_t\right) = e^{r(T+\Delta-t)}\left(2a_2K - a_1\right), \tag{2.69}$$

$$\begin{aligned}\dot{\rho}_t\left(k,k'\right) &= e^{r(T+\Delta-t)}\left[\dot{E}_t - \dot{H}_t\left(k+k'\right)\right] - r\left(1+\rho_t\left(k,k'\right)\right)\\ &= e^{r(T+\Delta-t)}\left[-a_1 + \left(k+k'\right)a_2 + \frac{2K-k-k'}{4}\frac{H_t^2}{\kappa\varepsilon_t + H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2 + \lambda_0\right]\right].\end{aligned}$$

This implies

$$\begin{aligned}\frac{d}{dt}\left[\rho_t\left(k,k'\right) - \varrho_t\right] &= e^{r(T+\Delta-t)}\left(2K-k-k'\right)\left[-a_2 + \frac{1}{4}\frac{H_t^2}{\kappa\varepsilon_t + H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2 + \lambda_0\right]\right]\\ &= -\left[\frac{a_2}{H_t} - \frac{1}{4}\frac{H_t}{\kappa\varepsilon_t + H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2 + \lambda_0\right]\right]\left[\rho_t\left(k,k'\right) - \varrho_t\right].\end{aligned}$$

Price dispersion. The standard deviation of the bilateral Federal funds rates is

$$\begin{aligned}\sigma_{\rho,t} &= \left\{\int\int\left[\rho_t\left(k,k'\right) - \varrho_t\right]^2 dF_t\left(k'\right)dF_t\left(k\right)\right\}^{1/2}\\ &= e^{r(T+\Delta-t)}H_t\left[\int\int\left(2K-k-k'\right)^2 dF_t\left(k'\right)dF_t\left(k\right)\right]^{1/2}\\ &= e^{r(T+\Delta-t)}H_t\left\{\int\int\left[\left(K-k\right)^2 + \left(K-k'\right)^2 + 2\left(K-k\right)\left(K-k'\right)\right]dF_t\left(k'\right)dF_t\left(k\right)\right\}^{1/2}\\ &= e^{r(T+\Delta-t)}H_t\cdot\sqrt{2}\sigma_{k,t}.\end{aligned}$$

This gives our measure of price dispersion.

Intermediation markup. By definition, the rate spread is

$$\begin{aligned}
& \Delta_{\rho,t}(k,q) \\
\equiv & \int \rho_t(k+q,k')\,dF_t(k') - \rho_t(k,q) \\
= & \int e^{r(T+\Delta-t)}\left[E_t - H_t(k+q+k')\right]dF_t(k') - e^{r(T+\Delta-t)}\left[E_t - H_t\left(k+k+\frac{2(\kappa\varepsilon_t+H_t)}{H_t}q\right)\right] \\
= & e^{r(T+\Delta-t)}\left[-H_t(k+q+K)+2kH_t+2(\kappa\varepsilon_t+H_t)q\right] \\
= & e^{r(T+\Delta-t)}\left[-H_t(K-k)+(2\kappa\varepsilon_t+H_t)q\right].
\end{aligned}$$

Thus the intermediation markup is given by taking $\Delta_{\rho,t}(k,q)$ differentiation with respect to q.

Unilization rate of trade opportunities. By definition,

$$\begin{aligned}
UR_t &= \frac{\int_k \int_{k'\geq k} m(\varepsilon_t,\varepsilon_t)\,q_t(k,k')\,dF_t(k')\,dF_t(k)}{TO_t} \\
&= \frac{\int_k \int_{k'\geq k} \frac{H_t(k'-k)}{2(\kappa\varepsilon_t+H_t)}\left[(\lambda-\lambda_0)\varepsilon_t^2+\lambda_0\right]dF_t(k')\,dF_t(k)}{TO_t} \\
&= \frac{H_t\left[(\lambda-\lambda_0)\varepsilon_t^2+\lambda_0\right]}{2(\kappa\varepsilon_t+H_t)}\frac{\int_k \int_{k'\geq k}(k'-k)\,dF_t(k')\,dF_t(k)}{TO_t} \\
&= \frac{H_t\left[(\lambda-\lambda_0)\varepsilon_t^2+\lambda_0\right]}{\kappa\varepsilon_t+H_t}.
\end{aligned}$$

Extensive margins. We provide a heuristic approach to derive the dynamics of the extensive margins. Let Δ be a small time length, and denote $m_t \equiv m(\varepsilon_t,\varepsilon_t)$ as the equilibrium matching rate. Then by definition,

$$1-P_t^{tr}(k) = (1-\Delta\cdot m_t)\left[1-P_{t+\Delta}^{tr}(k)\right]+\Delta\cdot m_t\cdot 0,$$

where $1-P_t^{tr}(k)$ denotes the probability of no trade over $[t,T]$ conditional on $k_t=k$, $1-\Delta\cdot m_t$ represents the probability of no meetings during $[t,t+\Delta]$, and 0 means the probability of

no trade is 0 given a meeting arrives at t. Take $\Delta \to 0$, we can obtain

$$\dot{P}_t^{tr}(k) = \lim_{\Delta \to 0} \frac{P_{t+\Delta}^{tr}(k) - P_t^{tr}(k)}{\Delta} = -m_t \left[1 - P_t^{tr}(k)\right].$$

The evolution of $P_t^b(k)$ and $P_t^s(k)$ can be derived similarly as follows.

$$1 - P_t^b(k) = (1 - \Delta \cdot m_t)\left[1 - P_{t+\Delta}^b(k)\right] + \Delta \cdot m_t \int_{k' \leq k} \left[1 - P_{t+\Delta}^b(k + q_t(k, k'))\right] dF_t(k'),$$

$$1 - P_t^s(k) = (1 - \Delta \cdot m_t)\left[1 - P_{t+\Delta}^s(k)\right] + \Delta \cdot m_t \int_{k' \geq k} \left[1 - P_{t+\Delta}^s(k + q_t(k, k'))\right] dF_t(k').$$

Take $\Delta \to 0$ gives

$$\dot{P}_t^b(k) = -m_t\left[1 - F_t(k)\right]\left[1 - P_t^b(k)\right] - m_t \int_{k' \leq k} \left[P_t^b(k + q_t(k, k')) - P_t^b(k)\right] dF_t(k'),$$

$$\dot{P}_t^s(k) = -m_t F_t(k)\left[1 - P_t^s(k)\right] - m_t \int_{k' \geq k} \left[P_t^s(k + q_t(k, k')) - P_t^s(k)\right] dF_t(k').$$

Then the evolution of $P_t^{int}(k)$ is

$$\begin{aligned}
\dot{P}_t^{int}(k) &= \dot{P}_t^b(k) + \dot{P}_t^s(k) - \dot{P}_t^{tr}(k) \\
&= -m_t \int_{k' \leq k} \left[P_t^b(k + q_t(k, k')) - P_t^{int}(k)\right] dF_t(k') \\
&\quad -m_t \int_{k' \geq k} \left[P_t^s(k + q_t(k, k')) - P_t^{int}(k)\right] dF_t(k').
\end{aligned}$$

Intensive margins. We provide an heuristic derivation of the absolute trades and net rades. First, for the individual absolute trades, let Δ be an infinitesimal time period. Then by the property of Poisson process,

$$\begin{aligned}
Q_t(k) &= \Delta \cdot m(\varepsilon_t, \varepsilon_t) \cdot \left[\int_{k'} |q_t(k, k')| \, dF_t(k') + \int_{k'} Q_{t+\Delta}(k + q_t(k, k')) \, dF_t(k')\right] \\
&\quad + \left[1 - \Delta \cdot m(\varepsilon_t, \varepsilon_t)\right] Q_{t+\Delta}(k).
\end{aligned}$$

Thus the aggregate absolute trades is given by

$$\begin{aligned} Q_t &= \int Q_t(k)\, dF_t(k) \\ &= \Delta \cdot m(\varepsilon_t, \varepsilon_t) \cdot \int_k \int_{k'} |q_t(k,k')|\, dF_t(k')\, dF_t(k) \\ &\quad + \int_k \left\{ \Delta \cdot m(\varepsilon_t, \varepsilon_t) \cdot \int_{k'} Q_{t+\Delta}(k + q_t(k,k'))\, dF_t(k') + [1 - \Delta \cdot m(\varepsilon_t, \varepsilon_t)]\, Q_{t+\Delta}(k) \right\} \\ &\quad \times dF_t(k) \\ &= \Delta \cdot m(\varepsilon_t, \varepsilon_t) \cdot \int_k \int_{k'} |q_t(k,k')|\, dF_t(k')\, dF_t(k) + Q_{t+\Delta}, \end{aligned}$$

where the last equality is given by the definition of $Q_t(k)$ and Q_t. Taking $\Delta \to 0$, we can obtain the following ODEs for $Q_t(k)$ and Q_t:

$$\begin{aligned} \dot{Q}_t(k) &= \lim_{\Delta \to 0} \frac{Q_{t+\Delta}(k) - Q_t(k)}{\Delta} \\ &= -m(\varepsilon_t, \varepsilon_t) \cdot \left[\int_{k'} |q_t(k,k')|\, dF_t(k') + \int_{k'} Q_{t+\Delta}(k + q_t(k,k'))\, dF_t(k') \right] \\ &\quad + m(\varepsilon_t, \varepsilon_t)\, Q_t(k), \end{aligned}$$

and

$$\begin{aligned} \dot{Q}_t &= \lim_{\Delta \to 0} \frac{Q_{t+\Delta} - Q_t}{\Delta} \\ &= -m(\varepsilon_t, \varepsilon_t) \cdot \int_k \int_{k'} |q_t(k,k')|\, dF_t(k')\, dF_t(k) \\ &= -m(\varepsilon_t, \varepsilon_t) \frac{H_t}{2(\kappa \varepsilon_t + H_t)} \int_k \int_{k'} |k' - k|\, dF_t(k')\, dF_t(k). \end{aligned}$$

This implies

$$Q = \int_0^T m(\varepsilon_t, \varepsilon_t) \frac{H_t}{2(\kappa \varepsilon_t + H_t)} \left(\int_k \int_{k'} |k' - k|\, dF_t(k')\, dF_t(k) \right) dt. \tag{2.70}$$

Second, for the individual net Federal funds purchase, it satisfies

$$\begin{aligned} L_t(k) &= \Delta \cdot m_t \int \frac{H_t(k'-k)}{2(\kappa\varepsilon_t + H_t)} dF_t(k') + \Delta \cdot m_t \int L_{t+\Delta}(k + q_t(k,k'))\, dF_t(k') \\ &\quad + (1 - \Delta \cdot m_t) L_{t+\Delta}(k) \\ &= \Delta \cdot m_t \frac{H_t(K-k)}{2(\kappa\varepsilon_t + H_t)} + \Delta \cdot m_t \int L_{t+\Delta}(k + q_t(k,k'))\, dF_t(k') \\ &\quad + (1 - \Delta \cdot m_t) L_{t+\Delta}(k). \end{aligned}$$

We guess and verify that $L_t(k) = \Theta_{1,t} - \Theta_{2,t}k$. Plugging the guessed formula into the above equation and matching the coefficients, we can get

$$\frac{\dot{\Theta}_{1,t}}{K} = \dot{\Theta}_{2,t} = \frac{m_t H_t}{2(\kappa\varepsilon_t + H_t)} (\Theta_{2,t} - 1).$$

With terminal condition $\Theta_{1,T} = \Theta_{2,T} = 0$, we have the following closed-form solution:

$$\begin{aligned} \Theta_{2,t} &= 1 - \exp\left[- \int_t^T \frac{m_z H_z}{2(\kappa\varepsilon_z + H_z)} dz \right], \\ \Theta_{1,t} &= K \cdot \Theta_{2,t}. \end{aligned}$$

Thus the individual net trades is given by

$$L_t(k) = \left\{ 1 - \exp\left[- \int_t^T \frac{m_z H_z}{2(\kappa\varepsilon_z + H_z)} dz \right] \right\} (K - k),$$

and the aggregate net trades is

$$L = \int |L_0(k)|\, dF_0(k) = \left\{ 1 - \exp\left[- \int_0^T \frac{m_z H_z}{2(\kappa\varepsilon_z + H_z)} dz \right] \right\} \int |K - k|\, dF_0(k). \quad (2.71)$$

Federal funds rate. The average Federal funds rate at t is given by equation (2.68). It satisfies the ODE (2.69) with terminal condition $1 + \varrho_T = e^{r\Delta}[A_1 - 2A_2 K]$, which has the

following closed-form solution:

$$\begin{aligned}
1+\varrho_t &= e^{r\Delta}\left(A_1-2A_2K\right)-\frac{2a_2K-a_1}{r}\left[e^{r(T+\Delta-t)}-e^{r\Delta}\right] \\
&= e^{r\Delta}\left[1+\gamma+\frac{\left(k_+-1\right)i^{DW}-\left(k_--1\right)i^{ER}}{k_+-k_-}\right]-\frac{2a_2K-a_1}{r}\left[e^{r(T+\Delta-t)}-e^{r\Delta}\right] \\
&= e^{r\Delta}\left[1+\gamma+i^{ER}+\frac{k_+-1}{k_+-k_-}\Delta i\right]-\frac{2a_2K-a_1}{r}\left[e^{r(T+\Delta-t)}-e^{r\Delta}\right].
\end{aligned}$$

2.C.13 Proof of Proposition 2.7

Proof.

Comparative statics for the length of search. Note that the length of search in this case is given by

$$\tau_2\left(\eta;A_2,T\right)=T+\frac{1}{r+\frac{\lambda_0}{4}}\log\left(1-\frac{\eta-A_2}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}\right). \tag{2.72}$$

Then the first column of table in the proposition is given by differentiating (2.72). That is,

$$\frac{\partial\tau_2\left(\eta;A_2,T\right)}{\partial i^{ER}}=\frac{1}{r+\frac{\lambda_0}{4}}\frac{1}{1-\frac{\eta-A_2}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}}\frac{\frac{a_2}{r+\frac{\lambda_0}{4}}-\eta}{\left(\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2\right)^2}\left[-\frac{1}{2K\left(k_+-k_-\right)}\right]<0,$$

$$\frac{\partial\tau_2\left(\eta;A_2,T\right)}{\partial i^{DW}}=\frac{1}{r+\frac{\lambda_0}{4}}\frac{1}{1-\frac{\eta-A_2}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}}\frac{\frac{a_2}{r+\frac{\lambda_0}{4}}-\eta}{\left(\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2\right)^2}\left[\frac{1}{2K\left(k_+-k_-\right)}\right]>0,$$

$$\frac{\partial\tau_2\left(\eta;A_2,T\right)}{\partial K}=\frac{1}{r+\frac{\lambda_0}{4}}\frac{1}{1-\frac{\eta-A_2}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}}\frac{\frac{a_2}{r+\frac{\lambda_0}{4}}-\eta}{\left(\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2\right)^2}\left[-\frac{i^{DW}-i^{ER}}{2K^2\left(k_+-k_-\right)}\right]<0,$$

$$\frac{\partial\tau_2\left(\eta;A_2,T\right)}{\partial\kappa}=\frac{1}{r+\frac{\lambda_0}{4}}\frac{1}{1-\frac{\eta-A_2}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}}\left\{-\frac{1}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}\left[\frac{\lambda}{2\left(\lambda-\lambda_0\right)}-1\right]\right\}<0,$$

$$\begin{aligned}\frac{\partial \tau_2(\eta; A_2, T)}{\partial \lambda_0} &= -\frac{1}{4\left(r+\frac{\lambda_0}{4}\right)^2}\log\left(1-\frac{\eta-A_2}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}\right) \\ &\quad +\frac{1}{r+\frac{\lambda_0}{4}}\frac{1}{1-\frac{\eta-A_2}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}}\left\{-\frac{\frac{\kappa\lambda}{2(\lambda-\lambda_0)^2}\left(\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2\right)+(\eta-A_2)\frac{a_2}{4\left(r+\frac{\lambda_0}{4}\right)^2}}{\left(\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2\right)^2}\right\} \\ &< 0,\end{aligned}$$

$$\frac{\partial \tau_2(\eta; A_2, T)}{\partial \lambda} = \frac{1}{r+\frac{\lambda_0}{4}}\frac{1}{1-\frac{\eta-A_2}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}}\frac{1}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}\frac{\kappa\lambda_0}{2(\lambda-\lambda_0)^2} > 0.$$

Comparative statics of $|q_t(k,k')|$. Note that $|q_t(k,k')| = \left|\frac{H_t(k'-k)}{2(\kappa\varepsilon_t+H_t)}\right|$, and $q_t(k,k') = \frac{k'-k}{2}$ for any $t > \tau_2(\eta; A_2, T)$. Thus we focus on the comparative statics over $t < \tau_2(\eta; A_2, T)$. The comparative statics with respect to i^{ER}, i^{DW} and K are given by differentiating H_t with respect to the terminal condition $H_T = A_2$. Note that H_t is monotonically decreasing in time in this case, and solving H_t backwards implies that increasing A_2 will shift the path of H_t upward. Since $\frac{\partial A_2}{\partial i^{ER}} < 0$, $\frac{\partial A_2}{\partial i^{DW}} > 0$ and $\frac{\partial A_2}{\partial K} < 0$, then we must have $\frac{\partial H_t}{\partial i^{ER}} < 0$, $\frac{\partial H_t}{\partial i^{DW}} > 0$ and $\frac{\partial H_t}{\partial K} < 0$, which gives the results in the table.

To obtain the comparative statics of $|q_t(k,k')|$ with respect to κ, denote $\tilde{q}_t \equiv \frac{H_t}{\kappa\varepsilon_t+H_t}$. When $\varepsilon_t = 1$,

$$\dot{\tilde{q}}_t = \tilde{q}_t(1-\tilde{q}_t)\left[r-\frac{a_2}{\kappa}\left(\frac{1}{\tilde{q}_t}-1\right)+\frac{\lambda}{4}\tilde{q}_t\right],$$

with $\dot{\tilde{q}}_t < 0$, $\frac{\partial \dot{\tilde{q}}_t}{\partial \kappa} > 0$. Moreover, $\varepsilon_t = 1$ iff $\tilde{q}_t \geq \frac{\eta}{\kappa+\eta} = \frac{2\lambda_0}{\lambda} - 1$. This implies that over $t \in [0, \tau_2(\eta; A_2, T)]$, the path of $\tilde{q}_t$ decreases slower and reaches $\frac{2\lambda_0}{\lambda} - 1$ under a larger κ. Therefore, for any $t < \tau_2(\eta; A_2, T)$, $|q_t(k,k')|$ decreases in κ.

For the comparative statics of $|q_t(k,k')|$ with respect to λ_0, note that $\frac{\partial \eta}{\partial \lambda_0} > 0$, $\frac{\partial \tau_2(\eta; A_2, T)}{\partial \lambda_0} < 0$, $\frac{\partial J(t; A_2, T)}{\partial \lambda_0} < 0$ and $\frac{\partial \dot{H}_t\big|_{\varepsilon_t=1}}{\partial H_t} > 0$. Moreover, for any $H_t \in [\eta, \mu_2]$, we have $\dot{H}_t\Big|_{\varepsilon_t=1} < \dot{H}_t\Big|_{\varepsilon_t=0}$. Therefore, when $t > \tau_2(\eta; A_2, T)$, we have $\frac{\partial H_t}{\partial \lambda_0} < 0$; when $t < \tau_2(\eta; A_2, T)$, H_t decreases slower and reaches a larger η. Solving the ODE of H_t backwards implies that H_t decreases

in λ_0 for any $t < \tau_2(\eta; A_2, T)$. Thus $|q_t(k, k')|$ also decreases in λ_0 for any $t < \tau_2(\eta; A_2, T)$.

For the comparative statics of $|q_t(k, k')|$ with respect to λ, note that when $t > \tau_2(\eta; A_2, T)$, ε_t, $q_t(k, k')$ and H_t are all independent of λ. So we focus on the comparative statics on $t < \tau_2(\eta; A_2, T)$. First, we show that H_t is concave in t on $t < \tau_2(\eta; A_2, T)$. To see this, note that when $\varepsilon_t = 1$, $\frac{\partial \dot{H}_t}{\partial H_t} > 0$ for any $H_t \geq 0$. Combining with that the equilibrium $\dot{H}_t < 0$ for any t, it follows that $\ddot{H}_t = \frac{\partial \dot{H}_t}{\partial H_t} \dot{H}_t < 0$ on $t < \tau_2(\eta; A_2, T)$.

Next, pick any $\lambda' > \lambda$ and we use x' to denote the value of an endogenous variable x under λ'. According to the previous results, we have $\tau_2'(\eta; A_2, T) > \tau_2(\eta; A_2, T)$ and $\mu_2' < \mu_2$. The concavity of H_t in t, and $\frac{\partial \dot{H}_t}{\partial \lambda} > 0$ guarantee that $H_t|_{\lambda'}$ interacts with $H_t|_{\lambda}$ at most once on $t < \tau_2'(\eta; A_2, T)$.

Next, simple algebra shows that $\dot{H}_{\tau_2'(\eta;A_2,T)-}\Big|_{\lambda'} < \dot{H}_{\tau_2'(\eta;A_2,T)-}\Big|_{\lambda}$. Since $H_{\tau_2'(\eta;A_2,T)}\big|_{\lambda'} = H_{\tau_2'(\eta;A_2,T)}\big|_{\lambda}$, it follows that $H_t|_{\lambda'} > H_t|_{\lambda}$ in the left neighborhood of $\tau_2'(\eta; A_2, T)$. Moreover, $\lim_{t\to-\infty} H_t|_{\lambda'} = \mu_2' < \mu_2 = \lim_{t\to-\infty} H_t|_{\lambda}$, thus there exists a unique point $\hat{t} < \tau_2'(\eta; A_2, T)$ such that $H_t|_{\lambda'} > H_t|_{\lambda}$ for $t \in \left(\hat{t}, \tau_2'(\eta; A_2, T)\right)$, and $H_t|_{\lambda'} < H_t|_{\lambda}$ for $t < \hat{t}$. Since the ODE of $\dot{H}_t$ is autonomous, then $\tau_2'(\eta; A_2, T) - \hat{t}$ is independent of T. Since $\tau_2'(\eta; A_2, T)$ is increasing in T, it follows that $\hat{t} < 0$ if and only if T is smaller than a threshold value $\hat{T}$. Given λ and λ', we can define $\hat{T}$ by setting $\hat{t} = 0$. The monotonicity of $\tau_2'(\eta; A_2, T)$ in T guarantees the uniqueness of $\hat{T}$. Therefore, when T is small enough, the bilateral trade size $|q_t(k, k')|$ is increasing in λ for $t < \tau_2(\eta; A_2, T)$.

Comparative statics of $L_0(k)$ and $\frac{\partial L_0(k)}{\partial k}$. The solution (2.38) to $L_t(k)$ implies that it suffices to show the comparative statics of $\Phi \triangleq \int_0^T \frac{m_t H_t}{2(\kappa\varepsilon_t + H_t)} dt$. Note that

$$\Phi = \int_0^T \frac{m_t H_t}{2(\kappa\varepsilon_t + H_t)} dt = \frac{\lambda}{2} \int_0^{\tau_2(\eta;A_2,T)} \frac{H_t}{\kappa + H_t} dt + \frac{\lambda_0}{2} \left[T - \tau_2(\eta; A_2, T)\right].$$

Thus

$$\begin{aligned}
\frac{\partial \Phi}{\partial A_2} &= \left(\frac{\lambda}{2}\frac{H_{\tau_2}}{\kappa + H_{\tau_2}} - \frac{\lambda_0}{2}\right)\frac{\partial \tau_2}{\partial A_2} + \frac{\lambda}{2}\int_0^{\tau_2} \frac{\kappa}{(\kappa + H_t)^2}\frac{\partial H_t}{\partial A_2}dt \\
&= -\frac{\lambda - \lambda_0}{2}\frac{\partial \tau_2}{\partial A_2} + \frac{\lambda}{2}\int_0^{\tau_2} \frac{\kappa}{(\kappa + H_t)^2}\left(-\frac{\partial \tau_2/\partial A_2}{\partial \tau_1\left(H_t;\eta,\tau_2\right)/\partial H_t}\right)dt \\
&= -\frac{\lambda - \lambda_0}{2}\frac{\partial \tau_2}{\partial A_2} + \frac{\lambda}{2}\frac{\partial \tau_2}{\partial A_2}\int_0^{\tau_2} \frac{\kappa}{(\kappa + H_t)^2}\left(-\dot{H}_t\right)dt \\
&= \left\{-\frac{\lambda - \lambda_0}{2} + \frac{\lambda}{2}\left[\frac{\kappa}{(\kappa + H_{\tau_2})} - \frac{\kappa}{(\kappa + H_0)}\right]\right\}\frac{\partial \tau_2}{\partial A_2} \\
&= \left(\frac{\lambda}{2}\frac{H_0}{H_0 + \kappa} - \frac{\lambda_0}{2}\right)\frac{\partial \tau_2}{\partial A_2}.
\end{aligned}$$

Since $\frac{\partial \tau_2}{\partial A_2} > 0$, then $\frac{\partial \Phi}{\partial A_2} < 0$ if and only if $H_0 < \frac{\lambda_0}{\lambda - \lambda_0}\kappa$. Since $H_T = A_2 < \eta < \frac{\lambda_0}{\lambda - \lambda_0}\kappa$, and H_t is decreasing over t, then $\frac{\partial \Phi}{\partial A_2} < 0$ if and only if T is sufficiently small. Since we assume a small T, we can get that

$$\frac{\partial L_0(k)}{\partial i^{ER}} \propto (K - k)\frac{\partial \Phi}{\partial A_2}\frac{\partial A_2}{\partial i^{ER}} \Rightarrow \operatorname{sgn}\left(\frac{\partial L_0(k)}{\partial i^{ER}}\right) = \operatorname{sgn}(K - k),$$

$$\frac{\partial L_0(k)}{\partial i^{DW}} \propto (K - k)\frac{\partial \Phi}{\partial A_2}\frac{\partial A_2}{\partial i^{DW}} \Rightarrow \operatorname{sgn}\left(\frac{\partial L_0(k)}{\partial i^{DW}}\right) = \operatorname{sgn}(k - K),$$

$$\frac{\partial^2 L_0(k)}{\partial k \partial i^{ER}} \propto -\frac{\partial \Phi}{\partial A_2}\frac{\partial A_2}{\partial i^{ER}} < 0,$$

$$\frac{\partial^2 L_0(k)}{\partial k \partial i^{DW}} \propto -\frac{\partial \Phi}{\partial A_2}\frac{\partial A_2}{\partial i^{DW}} > 0.$$

For the comparative statics of Φ w.r.t. κ, note that

$$\begin{aligned}
\frac{\partial \Phi}{\partial \kappa} &= \left(\frac{\lambda}{2}\frac{H_{\tau_2}}{\kappa + H_{\tau_2}} - \frac{\lambda_0}{2}\right)\frac{\partial \tau_2}{\partial \kappa} + \frac{\lambda}{2}\int_0^{\tau_2}\frac{\partial \tilde{q}_t}{\partial \kappa}dt \\
&= -\frac{\lambda - \lambda_0}{2}\frac{\partial \tau_2}{\partial \kappa} + \frac{\lambda}{2}\int_0^{\tau_2}\frac{\partial \tilde{q}_t}{\partial \kappa}dt.
\end{aligned}$$

The first term is positive and the second term is negative.Since τ_2 increases in T, then the

first term dominates under a small T. Thus when T is small,

$$\frac{\partial L_0(k)}{\partial \kappa} \propto (K-k)\frac{\partial \Phi}{\partial \kappa} \Rightarrow \operatorname{sgn}\left(\frac{\partial L_0(k)}{\partial \kappa}\right) = \operatorname{sgn}(K-k),$$

$$\frac{\partial^2 L_0(k)}{\partial k \partial \kappa} \propto -\frac{\partial \Phi}{\partial \kappa} < 0.$$

For the comparative statics of Φ w.r.t. λ_0, we have

$$\begin{aligned}\frac{\partial \Phi}{\partial \lambda_0} &= \left(\frac{\lambda}{2}\frac{H_{\tau_2}}{\kappa + H_{\tau_2}} - \frac{\lambda_0}{2}\right)\frac{\partial \tau_2}{\partial \lambda_0} + \frac{1}{2}\left[T - \tau_2(\eta; A_2, T)\right] + \frac{\lambda}{2}\int_0^{\tau_2}\frac{\partial \tilde{q}_t}{\partial \lambda_0}dt \\ &= -\frac{\lambda - \lambda_0}{2}\frac{\partial \tau_2}{\partial \lambda_0} + \frac{1}{2}\left[T - \tau_2(\eta; A_2, T)\right] + \frac{\lambda}{2}\int_0^{\tau_2}\frac{\partial \tilde{q}_t}{\partial \lambda_0}dt,\end{aligned}$$

where the first two terms are positive and the third term is negatve. Thus when T is small, we have $\frac{\partial \Phi}{\partial \lambda_0} > 0$, which implies

$$\frac{\partial L_0(k)}{\partial \lambda_0} \propto (K-k)\frac{\partial \Phi}{\partial \lambda_0} \Rightarrow \operatorname{sgn}\left(\frac{\partial L_0(k)}{\partial \lambda_0}\right) = \operatorname{sgn}(K-k),$$

$$\frac{\partial^2 L_0(k)}{\partial k \partial \lambda_0} \propto -\frac{\partial \Phi}{\partial \lambda_0} < 0.$$

For the comparative statics of Φ w.r.t. λ, we have

$$\begin{aligned}\frac{\partial \Phi}{\partial \lambda} &= \left(\frac{\lambda}{2}\frac{H_{\tau_2}}{\kappa + H_{\tau_2}} - \frac{\lambda_0}{2}\right)\frac{\partial \tau_2}{\partial \lambda} + \frac{1}{2}\int_0^{\tau_2}\frac{H_t}{\kappa + H_t}dt + \frac{\lambda}{2}\int_0^{\tau_2}\frac{\partial \tilde{q}_t}{\partial \lambda}dt \\ &= -\frac{\lambda - \lambda_0}{2}\frac{\partial \tau_2}{\partial \lambda} + \frac{1}{2}\int_0^{\tau_2}\frac{H_t}{\kappa + H_t}dt + \frac{\lambda}{2}\int_0^{\tau_2}\frac{\partial \tilde{q}_t}{\partial \lambda}dt,\end{aligned}$$

where the first term is negative and the last two terms are positive. When T is small, the first term dominates, thus we have $\frac{\partial \Phi}{\partial \lambda} < 0$. This implies

$$\frac{\partial L_0(k)}{\partial \lambda} \propto (K-k)\frac{\partial \Phi}{\partial \lambda} \Rightarrow \operatorname{sgn}\left(\frac{\partial L_0(k)}{\partial \lambda}\right) = \operatorname{sgn}(k-K),$$

$$\frac{\partial^2 L_0(k)}{\partial k \partial \lambda} \propto -\frac{\partial \Phi}{\partial \lambda} > 0.$$

For the comparative statics of Φ w.r.t. K, we have $\frac{\partial \Phi}{\partial K} = \frac{\partial \Phi}{\partial A_2}\frac{\partial A_2}{\partial K} > 0$, and

$$\frac{\partial L_0(k)}{\partial K} = 1 - \exp(-\Phi) + \exp(-\Phi)\frac{\partial \Phi}{\partial K}(K-k),$$

$$\frac{\partial^2 L_0(k)}{\partial k \partial K} \propto -\frac{\partial \Phi}{\partial A_2}\frac{\partial A_2}{\partial K} < 0.$$

This implies that

$$\frac{\partial L_0(k)}{\partial K}\begin{cases} < 0, & \text{if } k > K + \frac{\exp(\Phi)-1}{\partial \Phi / \partial K}, \\ > 0, & \text{otherwise.} \end{cases}$$

Comparative statics of $\rho_t(k,k')$. Using equation (2.21) and (2.22), we have

$$\begin{aligned} E_t &= e^{-r(T-t)}E_T + \int_t^T e^{-r(s-t)}\left\{a_1 - \frac{K}{2}\cdot\frac{H_s^2\left[(\lambda-\lambda_0)\varepsilon_s^2+\lambda_0\right]}{\kappa\varepsilon_s + H_s}\right\}ds, \\ H_t &= e^{-r(T-t)}H_T + \int_t^T e^{-r(s-t)}\left\{a_2 - \frac{1}{4}\cdot\frac{H_s^2\left[(\lambda-\lambda_0)\varepsilon_s^2+\lambda_0\right]}{\kappa\varepsilon_s + H_s}\right\}ds. \end{aligned}$$

Then we can write $\rho_t(k,k')$ as

$$\begin{aligned} \rho_t(k,k') &= e^{r(T+\Delta)}\left\{e^{-rT}\left[E_T - (k+k')H_T\right]\right. \\ &\quad \left. + \int_t^T e^{-rs}\left[a_1 - (k+k')a_2 + \frac{k+k'-2K}{4}\cdot\frac{H_s^2\left[(\lambda-\lambda_0)\varepsilon_s^2+\lambda_0\right]}{\kappa\varepsilon_s + H_s}\right]ds\right\}, \end{aligned} \tag{2.73}$$

where

$$E_T - (k+k')H_T = 1 + \frac{k_+ i^{DW} - k_- i^{ER}}{k_+ - k_-} + \gamma - \frac{i^{DW} - i^{ER}}{2K(k_+ - k_-)}(k+k'). \tag{2.74}$$

When $t > \tau_2$,

$$\begin{aligned}\int_t^T e^{-rs}\frac{\lambda_0}{4}\frac{\partial H_s}{\partial i^{DW}}ds &= \frac{\lambda_0}{4}\int_t^T e^{-rs-\left(r+\frac{\lambda_0}{4}\right)(T-s)}ds\frac{\partial A_2}{\partial i^{DW}} \\ &= e^{-rT}\left[1-e^{-\frac{\lambda_0}{4}(T-t)}\right]\frac{\partial A_2}{\partial i^{DW}} < e^{-rT}\frac{\partial A_2}{\partial i^{DW}}\end{aligned}$$

which implies that $\frac{\partial \rho_t(k,k')}{\partial i^{DW}} < 0$ iff

$$k + k' > 2K\frac{k_+ - 1 + \exp\left[-\frac{\lambda_0}{4}(T-t)\right]}{\exp\left[-\frac{\lambda_0}{4}(T-t)\right]}.$$

For $t < \tau_2$, note that $t = \tau_1\left(H_t; \eta, \tau_2\left(\eta; A_2, T\right)\right)$. The implicit function theorem implies that

$$\begin{aligned}\frac{\partial}{\partial i^{DW}}\left(\frac{\lambda H_t^2}{\kappa + H_t}\right) &= \lambda\frac{H_t\left(2\kappa + H_t\right)}{\left(\kappa + H_t\right)^2}\frac{\partial H_t}{\partial i^{DW}} \\ &= \lambda\frac{H_t\left(2\kappa + H_t\right)}{\left(\kappa + H_t\right)^2}\left(-\dot{H}_t\right)\frac{\partial \tau_2\left(\eta; A_2, T\right)}{\partial A_2}\frac{\partial A_2}{\partial i^{DW}}.\end{aligned}$$

Since the ODE of $\dot{H}_t$ is autonomous with $H_{\tau_2} = \eta$, it implies that for any $u > 0$, we can define the following $M\left(u\right)$:

$$M\left(u\right) = \frac{\lambda}{4}\frac{\partial \tau_2\left(\eta; A_2, T\right)}{\partial A_2}\int_{\tau_2(\eta;A_2,T)-u}^{\tau_2(\eta;A_2,T)} e^{-rs}\left[\frac{H_s\left(2\kappa + H_s\right)}{\left(\kappa + H_s\right)^2}\left(-\dot{H}_s\right)\right]ds.$$

which is independent of time t. Thus we can rewrite $\frac{\partial \rho_t(k,k')}{\partial i^{DW}}$ as

$$\frac{\partial \rho_t\left(k, k'\right)}{\partial i^{DW}} \propto e^{r(T+\Delta)}\left\{\begin{array}{l} e^{-rT}\frac{k_+ - (k+k')/2K}{k_+ - k_-} \\ +\frac{(k+k')/2K - 1}{k_+ - k_-}\left[M\left(\tau_2 - t\right) + e^{-rT}\left(1 - e^{-\frac{\lambda_0}{4}(T-\tau_2)}\right)\right]\end{array}\right\}.$$

It follows that $\frac{\partial \rho_t(k,k')}{\partial i^{DW}} < 0$ iff

$$k + k' > 2K\frac{k_+ - 1 + e^{-\frac{\lambda_0}{4}(T-\tau_2)} - M\left(\tau_2 - t\right)e^{rT}}{e^{-\frac{\lambda_0}{4}(T-\tau_2)} - M\left(\tau_2 - t\right)e^{rT}}.$$

Note that $M(0) = 0$ and $M(u)$ is increasing in u, then the above condition holds for sufficiently small $\tau_2 - t$. To guarantee the condition holds for all $t < \tau_2$, we need a sufficiently small T.

Similarly, the comparative statics of $\rho_t(k, k')$ w.r.t. i^{ER} is that when $t > \tau_2$, $\frac{\partial \rho_t(k,k')}{\partial i^{ER}} < 0$ iff

$$k + k' < 2K \frac{k_- - 1 + \exp\left[-\frac{\lambda_0}{4}(T-t)\right]}{\exp\left[-\frac{\lambda_0}{4}(T-t)\right]};$$

when $t < \tau_2$, $\frac{\partial \rho_t(k,k')}{\partial i^{ER}} < 0$ iff

$$k + k' < 2K \frac{k_- - 1 + e^{-\frac{\lambda_0}{4}(T-\tau_2)} - M(\tau_2 - t)\, e^{rT}}{e^{-\frac{\lambda_0}{4}(T-\tau_2)} - M(\tau_2 - t)\, e^{rT}}.$$

The comparative statics of $\rho_t(k, k')$ w.r.t. K is that when $t > \tau_2$, $\frac{\partial \rho_t(k,k')}{\partial K} < 0$ iff

$$k + k' < 2K \frac{\frac{\lambda_0}{4A_2}\int_t^T \exp\left[r(T-s)\right] H_s ds - 1 + \exp\left[-\frac{\lambda_0}{4}(T-t)\right]}{\exp\left[-\frac{\lambda_0}{4}(T-t)\right]};$$

when $t < \tau_2$, $\frac{\partial \rho_t(k,k')}{\partial i^{ER}} < 0$ iff

$$k + k' < 2K \frac{\int_t^T e^{r(T-s)} \frac{H_s^2\left[(\lambda-\lambda_0)\varepsilon_s^2 + \lambda_0\right]}{4A_2(\kappa\varepsilon_s + H_s)} ds - 1 + e^{-\frac{\lambda_0}{4}(T-\tau_2)} - M(\tau_2 - t)\, e^{rT}}{e^{-\frac{\lambda_0}{4}(T-\tau_2)} - M(\tau_2 - t)\, e^{rT}}.$$

The comparative statics of $\rho_t(k, k')$ w.r.t. λ, note that $\rho_t(k, k')$ is independent of λ on $t > \tau_2$. Thus we focus on $t < \tau_2$. In this case,

$$\begin{aligned}
\rho_t(k, k') &= e^{r(T+\Delta)} \left\{ e^{-rT} \left[E_T - (k + k') H_T \right] \right. \\
&\quad + \int_{\tau_2}^{T} e^{-rs} \left[a_1 - (k + k') a_2 + \frac{k + k' - 2K}{4} \cdot \lambda_0 H_s \right] ds \\
&\quad \left. + \int_{t}^{\tau_2} e^{-rs} \left[a_1 - (k + k') a_2 + \frac{k + k' - 2K}{4} \cdot \frac{\lambda H_s^2}{\kappa + H_s} \right] ds \right\},
\end{aligned}$$

and the derivative is

$$\begin{aligned}\frac{\partial \rho_t(k,k')}{\partial \lambda} &= e^{r(T+\Delta)}\left\{-e^{-r\tau_2}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\lambda_0\eta\right]\frac{\partial\tau_2}{\partial\lambda}\right.\\ &\quad +e^{-r\tau_2}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\frac{\lambda\eta^2}{\kappa+\eta}\right]\frac{\partial\tau_2}{\partial\lambda}\\ &\quad \left.+\int_t^{\tau_2}e^{-rs}\frac{k+k'-2K}{4}\cdot\frac{\partial}{\partial\lambda}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds\right\}\\ &= \frac{k+k'-2K}{4}e^{r(T+\Delta-\tau_2)}\left\{\int_t^{\tau_2}e^{r(\tau_2-s)}\frac{\partial}{\partial\lambda}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds-(\lambda-\lambda_0)\eta\frac{\partial\tau_2}{\partial\lambda}\right\}.\end{aligned}$$

Note that we have proved $\frac{\partial}{\partial\lambda}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]>0$, $\frac{\partial\tau_2}{\partial\lambda}>0$. Moreover, for $t<\tau_2$, the value of H_t only depends on η and the time difference τ_2-t. This means that $\int_t^{\tau_2}e^{r(\tau_2-s)}\frac{\partial}{\partial\lambda}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds$ is close to 0 for t close to τ_2. This implies that for T sufficiently small, $\int_t^{\tau_2}e^{r(\tau_2-s)}\frac{\partial}{\partial\lambda}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds-(\lambda-\lambda_0)\eta\frac{\partial\tau_2}{\partial\lambda}<0$ for any $t<\tau_2$. Then we can get that $\frac{\partial\rho_t(k,k')}{\partial\lambda}<(>)\,0$ iff $k+k'>(<)\,2K$.

The comparative statics of $\rho_t(k,k')$ w.r.t. κ is similar to λ. First, $\rho_t(k,k')$ is independent of λ on $t>\tau_2$. Thus we focus on $t<\tau_2$. In this case,

$$\begin{aligned}\frac{\partial \rho_t(k,k')}{\partial \kappa} &= e^{r(T+\Delta)}\left\{-e^{-r\tau_2}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\lambda_0\eta\right]\frac{\partial\tau_2}{\partial\kappa}\right.\\ &\quad +e^{-r\tau_2}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\frac{\lambda\eta^2}{\kappa+\eta}\right]\frac{\partial\tau_2}{\partial\kappa}\\ &\quad \left.+\int_t^{\tau_2}e^{-rs}\frac{k+k'-2K}{4}\cdot\frac{\partial}{\partial\kappa}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds\right\}\\ &= \frac{k+k'-2K}{4}e^{r(T+\Delta-\tau_2)}\left\{\int_t^{\tau_2}e^{r(\tau_2-s)}\frac{\partial}{\partial\kappa}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds-(\lambda-\lambda_0)\eta\frac{\partial\tau_2}{\partial\kappa}\right\}.\end{aligned}$$

Note that we have proved $\frac{\partial\tau_2}{\partial\kappa}<0$. Thus if T is sufficiently small, we have

$$\int_t^{\tau_2}e^{r(\tau_2-s)}\frac{\partial}{\partial\kappa}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds-(\lambda-\lambda_0)\eta\frac{\partial\tau_2}{\partial\kappa}>0$$

for any $t<\tau_2$. This implies that $\frac{\partial\rho_t(k,k')}{\partial\kappa}<(>)\,0$ iff $k+k'<(>)\,2K$.

The comparative statics of $\rho_t(k,k')$ w.r.t. λ_0 is as follows. First, when $t > \tau_2$, we have

$$\begin{aligned}\rho_t(k,k') &= e^{r(T+\Delta)}\left\{e^{-rT}\left[E_T-(k+k')H_T\right]\right.\\ &\quad\left.+\int_t^T e^{-rs}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\lambda_0 H_s\right]ds\right\},\end{aligned}$$

and

$$\frac{\partial\rho_t(k,k')}{\partial\lambda_0}=\frac{k+k'-2K}{4}e^{r(T+\Delta)}\int_t^T e^{-rs}\cdot\frac{\partial\left[\lambda_0 H_s\right]}{\partial\lambda_0}ds.$$

Note that $\frac{\partial[\lambda_0 H_t]}{\partial\lambda_0}=4\left[\frac{\partial\dot{H}_t}{\partial\lambda_0}-r\frac{\partial H_t}{\partial\lambda_0}\right]>0$, then we have $\frac{\partial\rho_t(k,k')}{\partial\lambda_0}<(>)\,0$ iff $k+k'<(>)\,2K$. Second, when $t<\tau_2$, we have

$$\begin{aligned}\frac{\partial\rho_t(k,k')}{\partial\lambda_0} &= e^{r(T+\Delta)}\left\{-e^{-r\tau_2}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\lambda_0\eta\right]\frac{\partial\tau_2}{\partial\lambda_0}\right.\\ &\quad+e^{-r\tau_2}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\frac{\lambda\eta^2}{\kappa+\eta}\right]\frac{\partial\tau_2}{\partial\lambda_0}\\ &\quad+\int_{\tau_2}^T e^{-rs}\frac{k+k'-2K}{4}\cdot\frac{\partial\left[\lambda_0 H_s\right]}{\partial\lambda_0}ds\\ &\quad\left.+\int_t^{\tau_2} e^{-rs}\frac{k+k'-2K}{4}\cdot\frac{\partial}{\partial\lambda_0}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds\right\}\\ &= \frac{k+k'-2K}{4}e^{r(T+\Delta-\tau_2)}\times\\ &\quad\left\{\int_{\tau_2}^T e^{-r(s-\tau_2)}\frac{\partial\left[\lambda_0 H_s\right]}{\partial\lambda_0}ds+\int_t^{\tau_2} e^{r(\tau_2-s)}\frac{\partial}{\partial\lambda_0}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds-(\lambda-\lambda_0)\eta\frac{\partial\tau_2}{\partial\lambda_0}\right\}.\end{aligned}$$

Since we have proved $\frac{\partial\tau_2}{\partial\lambda_0}<0$, $\frac{\partial[\lambda_0 H_s]}{\partial\lambda_0}>0$, then when T is small, the term in the big brackets is positive for any $t<\tau_2$. This implies that $\frac{\partial\rho_t(k,k')}{\partial\lambda_0}<(>)\,0$ iff $k+k'<(>)\,2K$. **Q.E.D.**

2.C.14 Proof of Proposition 2.8

Proof.

Comparative statics of the length of search. The length of search in this case is given by

$$T - \tau_1(\eta; A_2, T) = \frac{(\kappa + \mu_1)\log\left(\frac{A_2-\mu_1}{\eta-\mu_1}\right) - (\kappa+\mu_2)\log\left(\frac{A_2-\mu_2}{\eta-\mu_2}\right)}{\left(r+\frac{\lambda}{4}\right)(\mu_1-\mu_2)}. \tag{2.75}$$

Then the first column of table in the proposition is given by differentiating (2.75). Note that in this case, we have $\mu_1 < 0 < \mu_2 < \eta \leq A_2$. Therefore, we obtain

$$\frac{\partial\left[T-\tau_1(\eta;A_2,T)\right]}{\partial i^{ER}} = \frac{(\kappa+A_2)}{\left(r+\frac{\lambda}{4}\right)(A_2-\mu_1)(A_2-\mu_2)}\left[-\frac{1}{2K(k_+-k_-)}\right] < 0,$$

$$\frac{\partial\left[T-\tau_1(\eta;A_2,T)\right]}{\partial i^{DW}} = \frac{(\kappa+A_2)}{\left(r+\frac{\lambda}{4}\right)(A_2-\mu_1)(A_2-\mu_2)}\left[\frac{1}{2K(k_+-k_-)}\right] > 0,$$

$$\frac{\partial\left[T-\tau_1(\eta;A_2,T)\right]}{\partial K} = \frac{(\kappa+A_2)}{\left(r+\frac{\lambda}{4}\right)(A_2-\mu_1)(A_2-\mu_2)}\left[-\frac{i^{DW}-i^{ER}}{2K^2(k_+-k_-)}\right] < 0,$$

$$\frac{\partial\left[T-\tau_1(\eta;A_2,T)\right]}{\partial \lambda_0} = -\frac{\kappa+\eta}{(\eta-\mu_1)(\eta-\mu_2)\left(r+\frac{\lambda}{4}\right)}\frac{\kappa\lambda}{2(\lambda-\lambda_0)^2} < 0,$$

For the comparative statics w.r.t. κ, we define $\tilde{q}_t \equiv \frac{H_t}{\kappa+H_t}$. Following the derivations in 2.C.13, we have

$$\dot{\tilde{q}}_t = (1-\tilde{q}_t)\left[\frac{\lambda}{4}\tilde{q}_t^2 + \left(r+\frac{a_2}{\kappa}\right)\tilde{q}_t - \frac{a_2}{\kappa}\right],$$

with $\dot{\tilde{q}}_t > 0$, $\frac{\partial\dot{\tilde{q}}_t}{\partial\kappa} > 0$. Moreover, $\varepsilon_t = 1$ iff $\tilde{q}_t \geq \frac{\eta}{\kappa+\eta} = \frac{2\lambda_0}{\lambda} - 1$. This implies that over $t \in [\tau_1(\eta;A_2,T),T]$, when κ is larger, the $\tilde{q}_t$ increases faster from $\frac{2\lambda_0}{\lambda}-1$, and the terminal value $\frac{A_2}{A_2+\kappa}$ is smaller. Therefore, it takes less time for $\tilde{q}_t$ to increase from $\frac{2\lambda_0}{\lambda}-1$ to $\frac{A_2}{A_2+\kappa}$, i.e. $T-\tau_1(\eta;A_2,T)$ decreases in κ.

For the comparative statics w.r.t. λ, we define $\tilde{h}_t \equiv \lambda(\tilde{q}_t+1)$. Note that $\tilde{h}_t \in [2\lambda_0, 2\lambda]$. Then we have

$$\dot{\tilde{h}}_t = \left(2\lambda - \tilde{h}_t\right)\left\{\frac{1}{4\lambda}\tilde{h}_t^2 + \left[\frac{1}{\lambda}\left(r+\frac{a_2}{\kappa}\right) - \frac{1}{2}\right]\tilde{h}_t + \frac{\lambda}{4} - r - \frac{2a_2}{\kappa}\right\} > 0.$$

Moreover, we also have

$$\begin{aligned}\frac{\partial\dot{\tilde{h}}_t}{\partial\lambda} &= 2\left\{\frac{1}{4\lambda}\tilde{h}_t^2+\left[\frac{1}{\lambda}\left(r+\frac{a_2}{\kappa}\right)-\frac{1}{2}\right]\tilde{h}_t+\frac{\lambda}{4}-r-\frac{2a_2}{\kappa}\right\}\\ &\quad+\left(2\lambda-\tilde{h}_t\right)\left[\frac{1}{4}-\frac{1}{4\lambda^2}\tilde{h}_t^2-\frac{1}{\lambda^2}\left(r+\frac{a_2}{\kappa}\right)\tilde{h}_t\right],\end{aligned}$$

$$\frac{\partial^2\dot{\tilde{h}}_t}{\partial\lambda\partial\tilde{h}_t}=\frac{3}{4\lambda^2}\tilde{h}_t^2+\frac{2}{\lambda^2}\left(r+\frac{a_2}{\kappa}\right)\tilde{h}_t-\frac{5}{4},$$

with $\left.\frac{\partial^2\dot{\tilde{h}}_t}{\partial\lambda\partial\tilde{h}_t}\right|_{\tilde{h}_t=2\lambda}>0$. This implies that

$$\left.\frac{\partial\dot{\tilde{h}}_t}{\partial\lambda}\right|_{\tilde{h}_t=2\lambda}=2\left(r+\frac{\lambda}{4}\right)>0,$$

$$\left.\frac{\partial\dot{\tilde{h}}_t}{\partial\lambda}\right|_{\tilde{h}_t=\lambda}=-r-\frac{3a_2}{\kappa}<0,$$

and $\frac{\partial\dot{\tilde{h}}_t}{\partial\lambda}$ has a unique minimum on $\tilde{h}_t\in[\lambda,2\lambda]$, and is maximized at $\tilde{h}_t=2\lambda$. Since $2\lambda_0\in(\lambda,2\lambda)$, then $\left.\frac{\partial\dot{\tilde{h}}_t}{\partial\lambda}\right|_{\tilde{h}_t=2\lambda_0}<0$ iff $2\lambda_0$ is below a threshold point $\tilde{h}^*$. Note that $\tilde{h}_T=\lambda\left(\frac{A_2}{\kappa+A_2}+1\right)$ increases in λ, and $\tilde{h}_{\tau_1}=2\lambda_0$, then τ_1 decreases in λ if $\tilde{h}_{\tau_1}$ and $\tilde{h}_T$ are both below $\tilde{h}^*$. Therefore, $T-\tau_1(\eta;A_2,T)$ increases in λ if λ_0 and λ are both sufficiently small.

Comparative statics of $|q_t(k,k')|$. We focus on the comparative statics over $t>\tau_1$, during which q is variable. The comparative statics w.r.t. i^{ER}, i^{DW} and K are given by differentiating H_t w.r.t. A_2.Similar to the proof in 2.C.13, H_t increases in A_2. Then we must have $\frac{\partial|q_t(k,k')|}{\partial i^{ER}}<0$, $\frac{\partial|q_t(k,k')|}{\partial i^{DW}}>0$ and $\frac{\partial|q_t(k,k')|}{\partial K}<0$.

For the comparative statics w.r.t. κ, the proof for the length of search shows that over

$t \in [\tau_1(\eta; A_2, T), T]$, when κ is larger, the $\tilde{q}_t$ increases faster from $\frac{2\lambda_0}{\lambda} - 1$, and the terminal value $\frac{A_2}{A_2+\kappa}$ is smaller. Therefore,the path of $\tilde{q}_t$ over $[\tau_1(\eta; A_2, T), T]$ shifts downward under a larger κ, which implies $\frac{\partial |q_t(k,k')|}{\partial \kappa} < 0$.

For the comparative statics w.r.t. λ_0, note that H_t is independent of λ_0 on $t \in [\tau_1, T]$. Thus we have $\frac{\partial |q_t(k,k')|}{\partial \lambda_0} = 0$.

For the comparative statics w.r.t. λ, note that $\dot{H}_t$ increases in λ. This implies that as time goes from T to τ_1, H_t decreases faster from A_2 under a larger λ. Thus we must have $\frac{\partial |q_t(k,k')|}{\partial \lambda} < 0$.

Comparative statics of $L_0(k)$ and $\frac{\partial L_0(k)}{\partial k}$. Following the proof in 2.C.13, it suffices to show the comparative statics of Φ. Note that

$$\Phi = \int_0^T \frac{m_t H_t}{2(\kappa\varepsilon_t + H_t)} dt = \frac{\lambda}{2} \int_{\tau_1(\eta; A_2, T)}^T \frac{H_t}{\kappa + H_t} dt + \frac{\lambda_0}{2} \tau_1(\eta; A_2, T).$$

Thus

$$\begin{aligned}
\frac{\partial \Phi}{\partial A_2} &= \frac{\lambda - \lambda_0}{2} \frac{\partial \tau_1}{\partial A_2} - \frac{\lambda}{2} \frac{\partial \tau_1}{\partial A_2} \int_{\tau_1}^T \frac{\kappa}{(\kappa + H_t)^2} dH_t \\
&= \frac{\lambda - \lambda_0}{2} \frac{\partial \tau_1}{\partial A_2} - \frac{\lambda}{2} \frac{\partial \tau_1}{\partial A_2} \left(\frac{\kappa}{\kappa + \eta} - \frac{\kappa}{\kappa + A_2} \right) \\
&= \frac{\partial \tau_1}{\partial A_2} \frac{\kappa\lambda_0 - A_2(\lambda - \lambda_0)}{2(\kappa + A_2)}.
\end{aligned}$$

Since $\frac{\partial \tau_1}{\partial A_2} < 0$, then $\frac{\partial \Phi}{\partial A_2} > 0$ iff $\kappa < \frac{\lambda - \lambda_0}{\lambda_0} A_2$. Thus we assume a sufficiently small κ such that

$$\frac{\partial L_0(k)}{\partial i^{ER}} \propto (K - k) \frac{\partial \Phi}{\partial A_2} \frac{\partial A_2}{\partial i^{ER}} \Rightarrow \operatorname{sgn}\left(\frac{\partial L_0(k)}{\partial i^{ER}} \right) = \operatorname{sgn}(k - K)$$

$$\frac{\partial L_0(k)}{\partial i^{DW}} \propto (K - k) \frac{\partial \Phi}{\partial A_2} \frac{\partial A_2}{\partial i^{DW}} \Rightarrow \operatorname{sgn}\left(\frac{\partial L_0(k)}{\partial i^{DW}} \right) = \operatorname{sgn}(K - k),$$

$$\frac{\partial^2 L_0(k)}{\partial k \partial i^{ER}} \propto -\frac{\partial \Phi}{\partial A_2} \frac{\partial A_2}{\partial i^{ER}} > 0,$$

$$\frac{\partial^2 L_0(k)}{\partial k \partial i^{DW}} \propto -\frac{\partial \Phi}{\partial A_2}\frac{\partial A_2}{\partial i^{DW}} < 0.$$

For the comparative statics w.r.t. κ, we have

$$\frac{\partial \Phi}{\partial \kappa} = \frac{\lambda - \lambda_0}{2}\frac{\partial \tau_1}{\partial \kappa} + \frac{\lambda}{2}\int_{\tau_1}^{T} \frac{\partial \tilde{q}_t}{\partial \kappa} dt,$$

where the first term is positive and the second term is negative. Since we have assumed $\eta > 0$, it requires $\lambda \in (\lambda_0, 2\lambda_0)$. When $\lambda \to \lambda_0^+$, we have $\tau_1 \to T^-$ and

$$\lim_{\lambda \to \lambda_0^+} \frac{\partial \Phi}{\partial \kappa} = \lim_{\lambda \to \lambda_0^+} \frac{\lambda - \lambda_0}{2}\frac{\partial \tau_1}{\partial \kappa} = 0^+.$$

When $\lambda \to 2\lambda_0^-$, we have $\tau_1 \to 0^+$ and

$$\lim_{\lambda \to 2\lambda_0^-} \frac{\partial \Phi}{\partial \kappa} = \frac{\lambda}{2}\int_{0}^{T} \frac{\partial \tilde{q}_t}{\partial \kappa} dt < 0.$$

This implies that $\frac{\partial \Phi}{\partial \kappa} < 0$ when λ is sufficiently large relative to λ_0. As a consequence,

$$\frac{\partial L_0(k)}{\partial \kappa} \propto (K - k)\frac{\partial \Phi}{\partial \kappa} \Rightarrow \operatorname{sgn}\left(\frac{\partial L_0(k)}{\partial \kappa}\right) = \operatorname{sgn}(k - K),$$

$$\frac{\partial^2 L_0(k)}{\partial k \partial \kappa} \propto -\frac{\partial \Phi}{\partial \kappa} > 0.$$

For the comparative statics w.r.t. λ_0, we have

$$\frac{\partial \Phi}{\partial \lambda_0} = \frac{\lambda - \lambda_0}{2}\frac{\partial \tau_1}{\partial \lambda_0} + \frac{1}{2}\tau_1 > 0.$$

It follows that

$$\frac{\partial L_0(k)}{\partial \lambda_0} \propto (K - k)\frac{\partial \Phi}{\partial \lambda_0} \Rightarrow \operatorname{sgn}\left(\frac{\partial L_0(k)}{\partial \lambda_0}\right) = \operatorname{sgn}(K - k),$$

$$\frac{\partial^2 L_0(k)}{\partial k \partial \lambda_0} \propto -\frac{\partial \Phi}{\partial \lambda_0} < 0.$$

For the comparative statics of Φ w.r.t. λ, we define $\hat{h}_t \equiv \lambda \tilde{q}_t - \lambda_0$. Simple algebra reveals that on $t \in \hat{h}_t \in [\tau_1(\eta; A_2, T), T]$, $\hat{h}_t \in \left[\lambda_0 - \lambda, \lambda \frac{A_2}{\kappa + A_2} - \lambda_0\right] \subset (-\lambda_0, \lambda - \lambda_0)$, $\dot{\hat{h}}_t > 0$, $\left.\frac{\partial \dot{\hat{h}}_t}{\partial \lambda}\right|_{\hat{h}_t = -\lambda_0} = -\frac{a_2}{\kappa} < 0$, $\left.\frac{\partial \dot{\hat{h}}_t}{\partial \lambda}\right|_{\hat{h}_t = \lambda - \lambda_0} = r + \frac{\lambda}{4} > 0$, and $\frac{\partial \dot{\hat{h}}_t}{\partial \lambda}$ is negative (positive) if $\hat{h}_t$ is below (above) a threshold value $\hat{h}^* \in (-\lambda_0, \lambda - \lambda_0)$. It implies that when $\lambda_0 - \lambda$ and $\lambda \frac{A_2}{\kappa + A_2} - \lambda_0$ are both below $\hat{h}^*$, or equivalently λ_0/λ and λ are both sufficiently small, we must have that $\frac{\partial \hat{h}_t}{\partial \lambda} > 0$ and $\frac{\partial \tau_1}{\partial \lambda} < 0$. Moreover, we can write Φ as

$$\begin{aligned} \Phi &= \frac{\lambda}{2} \int_{\tau_1}^{T} \tilde{q}_t dt + \frac{\lambda_0}{2} \tau_1(\eta; A_2, T) \\ &= \frac{1}{2} \int_{\tau_1}^{T} (\lambda \tilde{q}_t - \lambda_0) dt + \frac{\lambda_0}{2} T = \frac{1}{2} \int_{\tau_1}^{T} \hat{h}_t dt + \frac{\lambda_0}{2} T. \end{aligned}$$

Given the above conditions on λ and λ_0, we can get $\frac{\partial \Phi}{\partial \lambda} > 0$. It implies

$$\frac{\partial L_0(k)}{\partial \lambda} \propto (K - k) \frac{\partial \Phi}{\partial \lambda} \Rightarrow \operatorname{sgn}\left(\frac{\partial L_0(k)}{\partial \lambda}\right) = \operatorname{sgn}(K - k),$$

$$\frac{\partial^2 L_0(k)}{\partial k \partial \lambda} \propto -\frac{\partial \Phi}{\partial \lambda} < 0.$$

For the comparative statics of Φ w.r.t. K, we have $\frac{\partial \Phi}{\partial K} = \frac{\partial \Phi}{\partial A_2} \frac{\partial A_2}{\partial K} < 0$, and

$$\frac{\partial L_0(k)}{\partial K} = 1 - \exp(-\Phi) + \exp(-\Phi) \frac{\partial \Phi}{\partial K}(K - k),$$

$$\frac{\partial^2 L_0(k)}{\partial k \partial K} \propto -\frac{\partial \Phi}{\partial A_2} \frac{\partial A_2}{\partial K} > 0.$$

This implies that

$$\frac{\partial L_0(k)}{\partial K} \begin{cases} < 0, & \text{if } k < K + \frac{\exp(\Phi) - 1}{\partial \Phi / \partial K}, \\ > 0, & \text{otherwise.} \end{cases}$$

Comparative statics of $\rho_t(k,k')$. The proof is similar to Section 2.C.13. For the comparative statics of $\rho_t(k,k')$ w.r.t. i^{DW}, we can take derivative to 2.73. When $t > \tau_1(\eta; A_2, T)$, we have

$$\begin{aligned}
\frac{\partial \rho_t(k,k')}{\partial i^{DW}} &= e^{r(T+\Delta)}\left\{e^{-rT}\left[\frac{k_+}{k_+-k_-}-\frac{k+k'}{2K(k_+-k_-)}\right]\right. \\
&\quad \left.+\frac{\lambda}{4}(k+k'-2K)\int_t^T e^{-rs}\frac{\partial}{\partial i^{DW}}\left[\frac{H_s^2}{\kappa+H_s}\right]ds\right\} \\
&= e^{r(T+\Delta)}\left\{e^{-rT}\left[\frac{k_+}{k_+-k_-}-\frac{k+k'}{2K(k_+-k_-)}\right]\right. \\
&\quad \left.+(k+k'-2K)\frac{\partial A_2}{\partial i^{DW}}\frac{\lambda}{4}\frac{\partial \tau_1(\eta; A_2, T)}{\partial A_2}\int_t^T e^{-rs}\frac{H_s(2\kappa+H_s)}{(\kappa+H_s)^2}(-dH_s)\right\} \\
&= e^{r(T+\Delta)}\left\{e^{-rT}\left[\frac{k_+}{k_+-k_-}-\frac{k+k'}{2K(k_+-k_-)}\right]\right. \\
&\quad \left.+\frac{(k+k'-2K)}{2K(k_+-k_-)}\tilde{M}(T-t)\right\},
\end{aligned}$$

where we define

$$\tilde{M}(u) \equiv \frac{\lambda}{4}\frac{\partial \tau_1(\eta; A_2, T)}{\partial A_2}\int_{T-u}^T e^{-rs}\frac{H_s(2\kappa+H_s)}{(\kappa+H_s)^2}(-dH_s),$$

which is positive and independent of t. Then we have $\frac{\partial \rho_t(k,k')}{\partial i^{DW}} < 0$ iff

$$k+k' > 2K\frac{k_+ - \tilde{M}(T-t)e^{rT}}{1-\tilde{M}(T-t)e^{rT}}.$$

When $t < \tau_1(\eta; A_2, T)$, we have

$$\begin{aligned}
\rho_t(k,k') &= e^{r(T+\Delta)}\left\{e^{-rT}\left[E_T-(k+k')H_T\right]\right. \\
&\quad +\int_{\tau_1}^T e^{-rs}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\frac{\lambda H_s^2}{\kappa+H_s}\right]ds \\
&\quad \left.+\int_t^{\tau_1} e^{-rs}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\lambda_0 H_s\right]ds\right\},
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \rho_t(k,k')}{\partial i^{DW}} &= e^{r(T+\Delta)}\left\{e^{-rT}\left[\frac{k_+}{k_+-k_-}-\frac{k+k'}{2K(k_+-k_-)}\right]+\frac{(k+k'-2K)}{2K(k_+-k_-)}\tilde{M}(T-\tau_1)\right. \\
&\quad -e^{-r\tau_1}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\frac{\lambda\eta^2}{\kappa+\eta}\right]\frac{\partial\tau_1}{\partial A_2}\frac{\partial A_2}{i^{DW}} \\
&\quad \left.+e^{-r\tau_1}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\lambda_0\eta\right]\frac{\partial\tau_1}{\partial A_2}\frac{\partial A_2}{i^{DW}}\right\} \\
&= e^{r(T+\Delta)}\left\{e^{-rT}\left[\frac{k_+}{k_+-k_-}-\frac{k+k'}{2K(k_+-k_-)}\right]+\frac{(k+k'-2K)}{2K(k_+-k_-)}\tilde{M}(T-\tau_1)\right. \\
&\quad \left.+\frac{k+k'-2K}{2K(k_+-k_-)}e^{-r\tau_1}\frac{(\lambda-\lambda_0)}{4}\eta\frac{\partial\tau_1}{\partial A_2}\right\}.
\end{aligned}$$

Thus we have $\frac{\partial\rho_t(k,k')}{\partial i^{DW}}<0$ iff

$$k+k'>2K\frac{k_+-\tilde{M}(T-\tau_1)e^{rT}-e^{r(T-\tau_1)}\frac{(\lambda-\lambda_0)\eta}{4}\frac{\partial\tau_1}{\partial A_2}}{1-\tilde{M}(T-\tau_1)e^{rT}-e^{r(T-\tau_1)}\frac{(\lambda-\lambda_0)\eta}{4}\frac{\partial\tau_1}{\partial A_2}}.$$

We can also derive the comparative statics w.r.t. i^{ER} in a similar way. The result is that when $t>\tau_1(\eta;A_2,T)$, $\frac{\partial\rho_t(k,k')}{\partial i^{ER}}<0$ iff

$$k+k'<2K\frac{k_--\tilde{M}(T-t)e^{rT}}{1-\tilde{M}(T-t)e^{rT}}.$$

When $t<\tau_1(\eta;A_2,T)$, $\frac{\partial\rho_t(k,k')}{\partial i^{ER}}<0$ iff

$$k+k'<2K\frac{k_--\tilde{M}(T-\tau_1)e^{rT}-e^{r(T-\tau_1)}\frac{(\lambda-\lambda_0)\eta}{4}\frac{\partial\tau_1}{\partial A_2}}{1-\tilde{M}(T-\tau_1)e^{rT}-e^{r(T-\tau_1)}\frac{(\lambda-\lambda_0)\eta}{4}\frac{\partial\tau_1}{\partial A_2}}.$$

For the comparative statics w.r.t. K, note that when $t > \tau_1$, we have

$$\begin{aligned}
\frac{\partial \rho_t(k,k')}{\partial K} &= e^{r(T+\Delta)}\left\{ e^{-rT}\frac{\left(i^{DW}-i^{ER}\right)(k+k')}{2K^2(k_+-k_-)} - \frac{\lambda}{2}\int_t^T e^{-rs}\frac{H_s^2}{\kappa+H_s}ds \right. \\
&\quad \left. +\frac{\lambda}{4}(k+k'-2K)\int_t^T e^{-rs}\frac{\partial}{\partial K}\left[\frac{H_s^2}{\kappa+H_s}\right]ds \right\} \\
&= e^{r(T+\Delta)}\left\{ e^{-rT}\frac{\left(i^{DW}-i^{ER}\right)(k+k')}{2K^2(k_+-k_-)} - \frac{\lambda}{2}\int_t^T e^{-rs}\frac{H_s^2}{\kappa+H_s}ds \right. \\
&\quad \left. -\frac{\left(i^{DW}-i^{ER}\right)(k+k'-2K)}{2K^2(k_+-k_-)}\tilde{M}(T-t) \right\}.
\end{aligned}$$

Thus $\frac{\partial \rho_t(k,k')}{\partial K} < 0$ iff

$$k+k' < 2K\frac{\frac{\lambda}{4A_2}\int_t^T e^{r(T-s)}\frac{H_s^2}{\kappa+H_s}ds - \tilde{M}(T-t)e^{rT}}{1-\tilde{M}(T-t)e^{rT}}.$$

When $t < \tau_1$, we have

$$\begin{aligned}
\frac{\partial \rho_t(k,k')}{\partial K} &= e^{r(T+\Delta)}\left\{ e^{-rT}\frac{\left(i^{DW}-i^{ER}\right)(k+k')}{2K^2(k_+-k_-)} - \frac{1}{2}\int_t^T e^{-rs}\frac{\left[(\lambda-\lambda_0)\varepsilon_s^2+\lambda_0\right]H_s^2}{\kappa\varepsilon_s+H_s}ds \right. \\
&\quad -\frac{\left(i^{DW}-i^{ER}\right)(k+k'-2K)}{2K^2(k_+-k_-)}\tilde{M}(T-\tau_1) \\
&\quad -e^{-r\tau_1}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\frac{\lambda\eta^2}{\kappa+\eta}\right]\frac{\partial\tau_1}{\partial A_2}\frac{\partial A_2}{K} \\
&\quad \left. +e^{-r\tau_1}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\lambda_0\eta\right]\frac{\partial\tau_1}{\partial A_2}\frac{\partial A_2}{K} \right\} \\
&= e^{r(T+\Delta)}\left\{ e^{-rT}\frac{\left(i^{DW}-i^{ER}\right)(k+k')}{2K^2(k_+-k_-)} - \frac{1}{2}\int_t^T e^{-rs}\frac{\left[(\lambda-\lambda_0)\varepsilon_s^2+\lambda_0\right]H_s^2}{\kappa\varepsilon_s+H_s}ds \right. \\
&\quad -\frac{\left(i^{DW}-i^{ER}\right)(k+k'-2K)}{2K^2(k_+-k_-)}\tilde{M}(T-\tau_1) \\
&\quad \left. -\frac{\left(i^{DW}-i^{ER}\right)(k+k'-2K)}{2K^2(k_+-k_-)}e^{-r\tau_1}\frac{(\lambda-\lambda_0)}{4}\eta\frac{\partial\tau_1}{\partial A_2} \right\}.
\end{aligned}$$

Thus $\frac{\partial \rho_t(k,k')}{\partial K} < 0$ iff

$$k + k' < 2K \frac{\int_{\tau_1}^{T} e^{r(T-s)} \frac{\left[(\lambda-\lambda_0)\varepsilon_s^2+\lambda_0\right]H_s^2}{4A_2(\kappa\varepsilon_s+H_s)} ds - \tilde{M}\,(T-t)\,e^{rT} - e^{r(T-\tau_1)}\frac{(\lambda-\lambda_0)}{4}\eta\frac{\partial\tau_1}{\partial A_2}}{1 - \tilde{M}\,(T-t)\,e^{rT} - e^{r(T-\tau_1)}\frac{(\lambda-\lambda_0)}{4}\eta\frac{\partial\tau_1}{\partial A_2}}.$$

For the comparative statics w.r.t. λ, we have that when $t > \tau_1$,

$$\begin{aligned}\rho_t\,(k,k') &= e^{r(T+\Delta)}\left\{e^{-rT}\left[E_T - (k+k')\,H_T\right]\right. \\ &\quad \left. + \int_t^T e^{-rs}\left[a_1 - (k+k')\,a_2 + \frac{k+k'-2K}{4}\cdot\frac{\lambda H_s^2}{\kappa+H_s}\right]ds,\right.\end{aligned}$$

and

$$\frac{\partial \rho_t\,(k,k')}{\partial\lambda} = \frac{k+k'-2K}{4} e^{r(T+\Delta)} \int_t^T e^{-rs}\cdot\frac{\partial}{\partial\lambda}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds.$$

Note that $\frac{\partial}{\partial\lambda}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right] = 4\left[\frac{\partial \dot{H}_t}{\partial\lambda} - r\frac{\partial H_t}{\partial\lambda}\right] > 0$. This implies that $\frac{\partial\rho_t(k,k')}{\partial\lambda} < 0$ iff $k + k' < 2K$. When $t < \tau_1$, we have

$$\begin{aligned}\rho_t\,(k,k') &= e^{r(T+\Delta)}\left\{e^{-rT}\left[E_T - (k+k')\,H_T\right]\right. \\ &\quad + \int_{\tau_1}^T e^{-rs}\left[a_1 - (k+k')\,a_2 + \frac{k+k'-2K}{4}\cdot\frac{\lambda H_s^2}{\kappa+H_s}\right]ds \\ &\quad \left. + \int_t^{\tau_1} e^{-rs}\left[a_1 - (k+k')\,a_2 + \frac{k+k'-2K}{4}\cdot\lambda_0 H_s\right]ds\right\},\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \rho_t(k,k')}{\partial \lambda} &= e^{r(T+\Delta)}\left\{-e^{-r\tau_1}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\frac{\lambda\eta^2}{\kappa+\eta}\right]\frac{\partial\tau_1}{\partial\lambda}\right. \\
&\quad +e^{-r\tau_1}\left[a_1-(k+k')a_2+\frac{k+k'-2K}{4}\cdot\lambda_0\eta\right]\frac{\partial\tau_1}{\partial\lambda} \\
&\quad +\int_{\tau_1}^{T}e^{-rs}\frac{k+k'-2K}{4}\cdot\frac{\partial}{\partial\lambda}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds \\
&\quad \left.+\int_{t}^{\tau_1}e^{-rs}\frac{k+k'-2K}{4}\cdot\frac{\partial\left[\lambda_0 H_s\right]}{\partial\lambda}ds\right\} \\
&= \frac{k+k'-2K}{4}e^{r(T+\Delta)}\left\{\int_{\tau_1}^{T}e^{-r(s-\tau_1)}\frac{\partial}{\partial\lambda}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds+\int_{t}^{\tau_1}e^{r(\tau_1-s)}\frac{\partial\left[\lambda_0 H_s\right]}{\partial\lambda}ds\right. \\
&\quad \left.+(\lambda-\lambda_0)\eta\frac{\partial\tau_1}{\partial\lambda}\right\},
\end{aligned}$$

where in the brackets the first term is positive and the last two terms are negative. When λ, λ_0 and T are sufficiently small, the first term dominates, and we have that $\frac{\partial\rho_t(k,k')}{\partial\lambda}<0$ iff $k+k'<2K$.

For the comparative statics w.r.t. κ, we have that when $t>\tau_1$,

$$\frac{\partial\rho_t(k,k')}{\partial\kappa}=\frac{k+k'-2K}{4}e^{r(T+\Delta)}\int_{t}^{T}e^{-rs}\cdot\frac{\partial}{\partial\kappa}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds.$$

Note that $\frac{\partial}{\partial\kappa}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]=4\left[\frac{\partial\dot{H}_t}{\partial\kappa}-r\frac{\partial H_t}{\partial\kappa}\right]<0$ due to $\frac{\partial\dot{H}_t}{\partial\kappa}<0$ and $\frac{\partial H_t}{\partial\kappa}>0$. This implies that

$\frac{\partial \rho_t(k,k')}{\partial \kappa} < 0$ iff $k + k' > 2K$. When $t < \tau_1$, we have

$$\begin{aligned}
\frac{\partial \rho_t(k,k')}{\partial \kappa} &= e^{r(T+\Delta)} \left\{ -e^{-r\tau_1} \left[a_1 - (k+k') a_2 + \frac{k+k'-2K}{4} \cdot \frac{\lambda \eta^2}{\kappa + \eta} \right] \frac{\partial \tau_1}{\partial \kappa} \right. \\
&\quad + e^{-r\tau_1} \left[a_1 - (k+k') a_2 + \frac{k+k'-2K}{4} \cdot \lambda_0 \eta \right] \frac{\partial \tau_1}{\partial \kappa} \\
&\quad + \int_{\tau_1}^{T} e^{-rs} \frac{k+k'-2K}{4} \cdot \frac{\partial}{\partial \kappa} \left[\frac{\lambda H_s^2}{\kappa + H_s} \right] ds \\
&\quad \left. + \int_{t}^{\tau_1} e^{-rs} \frac{k+k'-2K}{4} \cdot \frac{\partial [\lambda_0 H_s]}{\partial \kappa} ds \right\} \\
&= \frac{k+k'-2K}{4} e^{r(T+\Delta)} \left\{ \int_{\tau_1}^{T} e^{-r(s-\tau_1)} \frac{\partial}{\partial \kappa} \left[\frac{\lambda H_s^2}{\kappa + H_s} \right] ds + \int_{t}^{\tau_1} e^{r(\tau_1 - s)} \frac{\partial [\lambda_0 H_s]}{\partial \kappa} ds \right. \\
&\quad \left. + (\lambda - \lambda_0) \eta \frac{\partial \tau_1}{\partial \kappa} \right\},
\end{aligned}$$

where the first term is negative and the last two terms are positive. When λ, λ_0 and T are sufficiently small, the first term dominates, and we have that $\frac{\partial \rho_t(k,k')}{\partial \kappa} < 0$ iff $k + k' > 2K$.

For the comparative statics w.r.t. λ_0, we have that when $t > \tau_1$, $\rho_t(k,k')$ is independent of λ_0. When $t < \tau_1$, we have

$$\begin{aligned}
\frac{\partial \rho_t(k,k')}{\partial \lambda_0} &= e^{r(T+\Delta)} \left\{ -e^{-r\tau_1} \left[a_1 - (k+k') a_2 + \frac{k+k'-2K}{4} \cdot \frac{\lambda \eta^2}{\kappa + \eta} \right] \frac{\partial \tau_1}{\partial \lambda_0} \right. \\
&\quad + e^{-r\tau_1} \left[a_1 - (k+k') a_2 + \frac{k+k'-2K}{4} \cdot \lambda_0 \eta \right] \frac{\partial \tau_1}{\partial \lambda_0} \\
&\quad \left. + \int_{t}^{\tau_1} e^{-rs} \frac{k+k'-2K}{4} \cdot \frac{\partial [\lambda_0 H_s]}{\partial \lambda_0} ds \right\} \\
&= \frac{k+k'-2K}{4} e^{r(T+\Delta)} \left\{ \int_{t}^{\tau_1} e^{r(\tau_1 - s)} \frac{\partial [\lambda_0 H_s]}{\partial \lambda_0} ds + (\lambda - \lambda_0) \eta \frac{\partial \tau_1}{\partial \lambda_0} \right\}.
\end{aligned}$$

Note that $\frac{\partial [\lambda_0 H_t]}{\partial \lambda_0} = 4 \left[\frac{\partial \dot{H}_t}{\partial \lambda_0} - r \frac{\partial H_t}{\partial \lambda_0} \right] > 0$ and $\frac{\partial \tau_1}{\partial \lambda_0} > 0$, we have that $\frac{\partial \rho_t(k,k')}{\partial \lambda_0} < 0$ iff $k + k' < 2K$. **Q.E.D.**

2.C.15 Proof of Proposition 2.9

Proof. The proposition is a restatement of equations (2.12) and (2.14). **Q.E.D.**

2.C.16 Proof of Lemma 2.4

Proof. The proof follows Lemma 2.2. **Q.E.D.**

2.C.17 Proof of Proposition 2.10

Proof. To prove the inefficiencies on extensive margin, it is straightforward by showing that $\tau_1^p(\eta; A_2, T) > \tau_1(\eta; A_2, T)$ and $\tau_2^p(\eta; A, T) < \tau_2(\eta; A, T)$. To prove the inefficiencies on intensive margin, it suffices to show that $\dot{H}_t^p > \dot{H}_t$ for any $H_t^p = H_t$. To see this, note that the laws of motion of two variables can be written as

$$\begin{aligned}
\dot{H}_t &= rH_t - a_2 + \frac{1}{4}\frac{H_t^2\left[(\lambda-\lambda_0)\cdot 1\left\{H_t \geq \eta\right\} + \lambda_0\right]}{\kappa \cdot 1\left\{H_t \geq \eta\right\} + H_t}, \\
\dot{H}_t^p &= rH_t^p - a_2 + \frac{1}{2}\frac{\left(H_t^p\right)^2\left[(\lambda-\lambda_0)\cdot 1\left\{H_t \geq \eta^p\right\} + \lambda_0\right]}{\kappa \cdot 1\left\{H_t^p \geq \eta^p\right\} + H_t^p}.
\end{aligned}$$

Since $\eta = \eta^p$, we must have $\dot{H}_t^p > \dot{H}_t$ for any $H_t^p = H_t$. Then the terminal condition $H_T^p = H_T = A_2$ implies that $H_t^p < H_t$ for any t. Then the size of bilateral reallocation must satisfy

$$\left|q_t^p(k, k')\right| = \frac{H_t^p\left|k' - k\right|}{2\left(H_t^p + \kappa\right)} < \frac{H_t\left|k' - k\right|}{2\left(H_t + \kappa\right)} = \left|q_t(k, k')\right|$$

whenever there is active reallocation. **Q.E.D.**

2.D Appendix: Heterogeneous agents with peripheral traders

We guess and verify the closed-form solutions. First, we guess the banks' value function is $V_t(k) = -H_t k^2 + E_t k + D_t$, and the peripheral trader's value function is $\tilde{V}_t\left(\tilde{k}\right) = -\tilde{H}_t\tilde{k}^2 +$

$\tilde{E}_t\tilde{k} + \tilde{D}_t$. The terms of trade of a meeting between banks is similar to the baseline model, i.e.

$$S_t(k,k') = \frac{H_t^2(k'-k)^2}{\kappa_1(\varepsilon+\varepsilon') + 2\kappa_0 + 2H_t}, \tag{2.76}$$

$$q_t(k,k') = \frac{H_t(k'-k)}{\kappa_1(\varepsilon+\varepsilon') + 2\kappa_0 + 2H_t}. \tag{2.77}$$

The choice of optimal search intensity is given by equation (2.20). Thus the optimal search intensity of the most liquid equilibrium is given by

$$\varepsilon_t = \begin{cases} 1, & \text{if } H_t \geq \tilde{\eta} \equiv \kappa_1\left[\frac{\lambda}{2(\lambda-\lambda_0)} - 1\right] - \kappa_0, \\ 0, & \text{otherwise.} \end{cases}$$

For the meetings between a bank and a peripheral trader, they solve

$$\max_{R,q}\left[V_t(k+q) - e^{-r(T-t+\Delta)}R - V_t(k) - \chi(0,q)\right]^{\theta} \times \left[\tilde{V}_t\left(\tilde{k}-q\right) + e^{-r(T-t+\Delta)}R - \tilde{V}_t\left(\tilde{k}\right)\right]^{1-\theta}.$$

The maximized surplus and optimal trade size are given by

$$\tilde{S}_t\left(k,\tilde{k}\right) = \frac{\left[E_t - \tilde{E}_t + 2\left(\tilde{H}_t\tilde{k} - H_t k\right)\right]^2}{4\left[H_t + \tilde{H}_t + \kappa_0\right]}, \tag{2.78}$$

$$\tilde{q}_t\left(k,\tilde{k}\right) = \frac{E_t - \tilde{E}_t + 2\left(\tilde{H}_t\tilde{k} - H_t k\right)}{2\left[H_t + \tilde{H}_t + \kappa_0\right]}. \tag{2.79}$$

Therefore, the HJB for peripheral traders is

$$r\tilde{V}_t\left(\tilde{k}\right) = \dot{\tilde{V}}_t\left(\tilde{k}\right) + (1-\theta)\varphi\int \tilde{S}_t\left(k,\tilde{k}\right)dF_t(k).$$

By matching coefficients we can obtain

$$\begin{aligned}
\dot{\tilde{H}}_t &= r\tilde{H}_t + \frac{(1-\theta)\,\varphi\tilde{H}_t^2}{H_t+\tilde{H}_t+\kappa_0},\ \text{with } \tilde{H}_T = 0;\\
\dot{\tilde{E}}_t &= r\tilde{E}_t - (1-\theta)\,\varphi\tilde{H}_t\frac{E_t-\tilde{E}_t-2H_tK_t}{H_t+\tilde{H}_t+\kappa_0},\ \text{with } \tilde{E}_T = 1+i^{RRP};\\
\dot{\tilde{D}}_t &= r\tilde{D}_t - (1-\theta)\,\varphi\int\frac{\left[E_t-\tilde{E}_t-2H_tk\right]^2}{4\left[H_t+\tilde{H}_t+\kappa_0\right]}dF_t(k),\ \text{with } \tilde{D}_t = 0.
\end{aligned}$$

Given $\tilde{H}_T = 0$, $\tilde{E}_T = 1+i^{RRP}$, we can get that $\tilde{H}_t \equiv 0$ and $\tilde{E}_t = \left(1+i^{RRP}\right)e^{-r(T-t)}$. Thus the bilateral Federal funds rate in a meeting between bank and peripheral trader is

$$\begin{aligned}
1+\tilde{\rho}_t\left(k,\tilde{k}\right) &= e^{r(T+\Delta-t)}\left[\frac{1-\theta}{2}\left(E_t-\tilde{E}_t-2H_tk\right)+\tilde{E}_t\right]\\
&= e^{r(T+\Delta-t)}\left[(1-\theta)\,(H_t+\kappa_0)\,\tilde{q}_t\left(k,\tilde{k}\right)+\tilde{E}_t\right]
\end{aligned}$$

On the other hand, the HJB for banks is

$$rV_t(k) = \dot{V}_t(k) + u(k) + \int\frac{1}{2}S_t(k,k')\,m(\varepsilon_t,\varepsilon_t)\,dF_t(k') + \theta\varphi\vartheta\int\tilde{S}_t\left(k,\tilde{k}\right)d\tilde{F}_t\left(\tilde{k}\right),$$

which implies

$$\dot{H}_t = rH_t - a_2 + \frac{1}{4}\frac{H_t^2}{\kappa_1\varepsilon_t+\kappa_0+H_t}\left[(\lambda-\lambda_0)\,\varepsilon_t^2+\lambda_0\right] + \frac{\theta\varphi\vartheta H_t^2}{H_t+\kappa_0}, \tag{2.80}$$

$$\dot{E}_t = rE_t - a_1 + \frac{K_t}{2}\frac{H_t^2}{\kappa_1\varepsilon_t+\kappa_0+H_t}\left[(\lambda-\lambda_0)\,\varepsilon_t^2+\lambda_0\right] + \theta\varphi\vartheta H_t\frac{E_t-\tilde{E}_t}{H_t+\kappa_0}, \tag{2.81}$$

$$D_t = rD_t - \frac{1}{4}\frac{H_t^2}{\kappa_1\varepsilon_t+\kappa_0+H_t}\left[(\lambda-\lambda_0)\,\varepsilon_t^2+\lambda_0\right]\int k'^2dF_t(k') - \theta\varphi\vartheta\frac{\left[E_t-\tilde{E}_t\right]^2}{4\left[H_t+\kappa_0\right]} \tag{2.82}$$

It follows that

$$\frac{d\left(E_t-\tilde{E}_t\right)}{dt}=\left(r+\frac{\theta\rho\varphi H_t}{H_t+\kappa_0}\right)\left(E_t-\tilde{E}_t\right)-a_1+\frac{K_t}{2}\frac{H_t^2}{\kappa_1\varepsilon_t+\kappa_0+H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2+\lambda_0\right], \tag{2.83}$$

and

$$\dot{K}_t=\varphi\vartheta\int\frac{E_t-\tilde{E}_t+2\left(\tilde{H}_t\tilde{k}-H_tk\right)}{2\left[H_t+\tilde{H}_t+\kappa_0\right]}dF_t\left(k\right)=\varphi\vartheta\frac{E_t-\tilde{E}_t-2H_tK_t}{2\left[H_t+\kappa_0\right]}, \tag{2.84}$$

with the boundary condition $K_0=K$ and $E_T-\tilde{E}_T=A_1-1-i^{RRP}$. We focus on the numerical solution.

The most liquid equilibrium. To characterize the dynamics of the most liquid equilibrium, we first define ω_1, ω_2 and ω_3 as the three real roots of H to the equation (the three real roots must exist by graphic proof)

$$\begin{aligned}0 &= \left(r+\frac{\lambda}{4}+\theta\varphi\vartheta\right)H^3+\left[r\left(2\kappa_0+\kappa_1\right)-a_2+\frac{\kappa_0\lambda}{4}+\theta\varphi\vartheta\left(\kappa_0+\kappa_1\right)\right]H^2\\ &\quad+\left[r\kappa_0\left(\kappa_0+\kappa_1\right)-a_2\left(2\kappa_0+\kappa_1\right)\right]H-a_2\kappa_0\left(\kappa_0+\kappa_1\right).\end{aligned}$$

Let $A\equiv r+\frac{\lambda}{4}+\theta\varphi\vartheta$, $B\equiv r\left(2\kappa_0+\kappa_1\right)-a_2+\frac{\kappa_0\lambda}{4}+\theta\varphi\vartheta\left(\kappa_0+\kappa_1\right)$, $C\equiv r\kappa_0\left(\kappa_0+\kappa_1\right)-a_2\left(2\kappa_0+\kappa_1\right)$ and $D\equiv-a_2\kappa_0\left(\kappa_0+\kappa_1\right)$, then the solution to ω_1, ω_2 and ω_3 are given by

$$\begin{aligned}\omega_1 &= \frac{-B}{3A}+\sqrt[3]{\frac{BC}{6A^2}-\frac{B^3}{27A^3}-\frac{D}{2A}+\sqrt{\left(\frac{BC}{6A^2}-\frac{B^3}{27A^3}-\frac{D}{2A}\right)^2+\left(\frac{C}{3A}-\frac{B^2}{9A^2}\right)^3}}\\ &\quad+\sqrt[3]{\frac{BC}{6A^2}-\frac{B^3}{27A^3}-\frac{D}{2A}-\sqrt{\left(\frac{BC}{6A^2}-\frac{B^3}{27A^3}-\frac{D}{2A}\right)^2+\left(\frac{C}{3A}-\frac{B^2}{9A^2}\right)^3}},\end{aligned}$$

$$\begin{aligned}\omega_2 &= \frac{-B}{3A}+\frac{-1+\sqrt{3}i}{2}\sqrt[3]{\frac{BC}{6A^2}-\frac{B^3}{27A^3}-\frac{D}{2A}+\sqrt{\left(\frac{BC}{6A^2}-\frac{B^3}{27A^3}-\frac{D}{2A}\right)^2+\left(\frac{C}{3A}-\frac{B^2}{9A^2}\right)^3}} \\ &\quad +\frac{-1-\sqrt{3}i}{2}\sqrt[3]{\frac{BC}{6A^2}-\frac{B^3}{27A^3}-\frac{D}{2A}-\sqrt{\left(\frac{BC}{6A^2}-\frac{B^3}{27A^3}-\frac{D}{2A}\right)^2+\left(\frac{C}{3A}-\frac{B^2}{9A^2}\right)^3}},\end{aligned}$$

$$\begin{aligned}\omega_3 &= \frac{-B}{3A}+\frac{-1-\sqrt{3}i}{2}\sqrt[3]{\frac{BC}{6A^2}-\frac{B^3}{27A^3}-\frac{D}{2A}+\sqrt{\left(\frac{BC}{6A^2}-\frac{B^3}{27A^3}-\frac{D}{2A}\right)^2+\left(\frac{C}{3A}-\frac{B^2}{9A^2}\right)^3}} \\ &\quad +\frac{-1+\sqrt{3}i}{2}\sqrt[3]{\frac{BC}{6A^2}-\frac{B^3}{27A^3}-\frac{D}{2A}-\sqrt{\left(\frac{BC}{6A^2}-\frac{B^3}{27A^3}-\frac{D}{2A}\right)^2+\left(\frac{C}{3A}-\frac{B^2}{9A^2}\right)^3}}.\end{aligned}$$

Next, denote β_1, β_2 and β_3 as the solution to the follow linear equation system:

$$\begin{bmatrix} 1 & 1 & 1 \\ \omega_2+\omega_3 & \omega_1+\omega_3 & \omega_1+\omega_2 \\ \omega_2\omega_3 & \omega_1\omega_3 & \omega_1\omega_2 \end{bmatrix}\begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\end{bmatrix}=\begin{bmatrix}1\\ -(2\kappa_0+\kappa_1)\\ \kappa_0(\kappa_1+\kappa_0)\end{bmatrix},$$

and define

$$\begin{aligned}\tilde{\mu}_1 &\equiv \frac{1}{2r+\frac{\lambda_0}{2}+2\theta\rho\varphi}\left\{-(\kappa_0 r-a_2)-\left[(\kappa_0 r-a_2)^2+a_2\kappa_0(4r+\lambda_0+4\theta\rho\varphi)\right]^{0.5}\right\}, \\ \tilde{\mu}_2 &\equiv \frac{1}{2r+\frac{\lambda_0}{2}+2\theta\rho\varphi}\left\{-(\kappa_0 r-a_2)+\left[(\kappa_0 r-a_2)^2+a_2\kappa_0(4r+\lambda_0+4\theta\rho\varphi)\right]^{0.5}\right\},\end{aligned}$$

$$\tilde{\tau}_1(H;A,u)\equiv u-\frac{1}{r+\frac{\lambda}{4}+\theta\rho\varphi}\left[\beta_1\log\left(\frac{A-\omega_1}{H-\omega_1}\right)+\beta_2\log\left(\frac{A-\omega_2}{H-\omega_2}\right)+\beta_3\log\left(\frac{A-\omega_3}{H-\omega_3}\right)\right],$$

and

$$\tilde{\tau}_2(H;A,u)\equiv u-\frac{(\kappa_0+\tilde{\mu}_1)\log\left(\frac{A-\tilde{\mu}_1}{H-\tilde{\mu}_1}\right)-(\kappa_0+\tilde{\mu}_2)\log\left(\frac{A-\tilde{\mu}_2}{H-\tilde{\mu}_2}\right)}{\left(r+\frac{\lambda_0}{4}+\theta\rho\varphi\right)(\tilde{\mu}_1-\tilde{\mu}_2)}.$$

Then the following proposition characterizes the path of equilibrium search prfile in the most liquid equilibrium.

Proposition 2.12 *(a). Suppose* $A_2 \geq \tilde{\eta}$.

(a-i). If $\dot{H}_t\Big|_{\varepsilon_t=1, H_t=\tilde{\eta}} > 0$ *and* $\tilde{\tau}_1(\tilde{\eta}; A_2, T) > 0$, *then we have*

$$\varepsilon_t = \begin{cases} 1, & \text{if } t \geq \tilde{\tau}_1(\tilde{\eta}; A_2, T); \\ 0, & \text{otherwise.} \end{cases}$$

$$H_t = \begin{cases} \tilde{\tau}_1^{-1}(t; A_2, T), & \text{if } t \geq \tilde{\tau}_1(\tilde{\eta}; A_2, T); \\ \tilde{\tau}_2^{-1}(t; \tilde{\eta}, \tilde{\tau}_1(\tilde{\eta}; A_2, T)), & \text{otherwise.} \end{cases}$$

(a-ii). Otherwise, we have $\varepsilon_t = 1$ *for all* $t \in [0, T]$ *and* $H_t = \tilde{\tau}_1^{-1}(t; A_2, T)$.

(b). Suppose $A_2 < \tilde{\eta}$.

(b-i). If $\dot{H}_t\Big|_{\varepsilon_t=0, H_t=\tilde{\eta}} < 0$ *and* $\tilde{\tau}_2(\tilde{\eta}; A_2, T) > 0$, *then we have*

$$\varepsilon_t = \begin{cases} 0, & \text{if } t > \tilde{\tau}_2(\tilde{\eta}; A_2, T); \\ 1, & \text{otherwise.} \end{cases}$$

$$H_t = \begin{cases} \tilde{\tau}_2^{-1}(t; A_2, T), & \text{if } t \geq \tilde{\tau}_2(\tilde{\eta}; A_2, T); \\ \tilde{\tau}_1^{-1}(t; \tilde{\eta}, \tilde{\tau}_2(\tilde{\eta}; A_2, T)), & \text{otherwise.} \end{cases}$$

(b-ii). Otherwise, we have $\varepsilon_t = 0$ *for all* $t \in [0, T]$ *and* $H_t = \tilde{\tau}_2^{-1}(t; A_2, T)$.

Proof. The proof follows Lemma 2.2. **Q.E.D.**

2.E Appendix: Federal funds brokerage

In this section we model the brokerage of Federal funds following Lagos and Rocheteau (2007). In practice, Federal funds brokers reach out their banks' contact for matchmaking. Consider the following timing of actions. Having secured a pair of banks for potential Federal funds trading, the broker negotiates with each banks about its brokerage fee. In this stage,

the broker does not reveal the identities of counterparties but informs the banks about the reserve balances held by their counterparties (the sufficient information banks need to know to initiate Federal funds trade in this model). This prevents the side-trading between the counterparty banks circumventing the broker's fee. Having determined the brokerage fees, the identities are revealed and the two banks negotitate the terms of trade like any bilateral Federal funds trades we described before. The brokerage fee is settled in numéraire at $T+\Delta$. We assume the matching rate between a broker and the bank counterparties is α, thus the contact rate of banks with a broker is $\alpha\nu$, where ν is the measure of active brokers. Brokers are free entry with entry cost ψ per broker.

We solve the outcome backward. Consider that a broker has identified a k-bank and a k'-bank at t. Each bank anticipates their trade surplus from trading with the arranged counterparties as $0.5S_t(k,k')$. Denote $Y_t(k,k')$ as the brokerage fee paid by k-bank for arranging the match with k'-bank; vice versa for the brokerage fee $Y_t(k',k)$ paid by k'-bank. To the k-bank, the surplus of brokerage is $0.5S_t(k,k') - Y_t(k,k')$. To the broker, the surplus of brokeraging the side of k-bank is simply $Y_t(k,k')$. Thus, the brokerage fee solves the following Nash bargaining problem:

$$Y_t(k,k') = \arg\max_y \{y[0.5S_t(k,k') - y]\}.$$

Hence the bargaining solution is

$$Y_t(k,k') = Y_t(k',k) = 0.25S_t(k,k').$$

The value of the broker, J_t, solves the following HJB equation

$$rJ_t = \dot{J}_t + \alpha \int\int [Y_t(k,k') + Y_t(k',k)]\, dF_t(k')\, dF_t(k), \text{ where } J_T = 0.$$

Denote the dependence of J_t on the broker size ν as $J_t(\nu)$. In the equilibrium, ν is determined

by the free-entry condition to the brokers:

$$\psi = J_0(\nu).$$

The bank's HJB is

$$\begin{aligned} rV_t(k) &= \dot{V}_t(k) + u(k) + \max_{\varepsilon_t \in [0,1]} \int \frac{1}{2} S_t(k, k', \varepsilon_t, \varepsilon_t(k'))\, m(\varepsilon_t, \varepsilon_t(k'))\, dF_t(k') \\ &\quad + \alpha\nu \int \frac{1}{4} S_t(k, k', 0, 0)\, dF_t(k') \end{aligned}$$

With quadratic utility function, we guess and verify $V_t(k) = -H_t k^2 + E_t k + D_t$, the solution is

$$\begin{aligned} rV_t(k) &= \dot{V}_t(k) + u(k) + \frac{1}{2} \frac{(H_t)^2}{\kappa(\varepsilon + \varepsilon_t) + 2H_t} \left[(\lambda - \lambda_0)\varepsilon\varepsilon_t + \lambda_0\right] \int (k' - k)^2 dF_t(k') \\ &\quad + \frac{\alpha\nu}{8} H_t \int (k' - k)^2 dF_t(k') \end{aligned}$$

By matching coefficients we obtain

$$\left(r + \frac{\alpha\nu}{8}\right) H_t = \dot{H}_t + a_2 - \frac{1}{4} \frac{H_t^2}{\kappa\varepsilon_t + H_t} \left[(\lambda - \lambda_0)\varepsilon_t^2 + \lambda_0\right], \tag{2.85}$$

thus $\alpha\nu$ changes the discount rate to the banks. The surplus function is

$$S_t(k, k', \varepsilon_t) = \frac{\left[H_t(k' - k)\right]^2}{2(\kappa\varepsilon_t + H_t)}$$

and the broker's HJB is

$$\begin{aligned} rJ_t &= \dot{J}_t + \frac{\alpha}{2} \int \int S_t(k, k', 0)\, dF_t(k')\, dF_t(k) \qquad (2.86) \\ &= \dot{J}_t + \frac{\alpha}{4} H_t \left[\int k^2 dF_t(k) - K^2\right] \end{aligned}$$

The solution is

$$J_0(\nu) = \frac{\alpha}{4}\int_0^T e^{-rt} H_t \left[\int k^2 dF_t(k) - K^2\right] dt, \tag{2.87}$$

where

$$\begin{aligned} &\int k^2 dF_t(k) \\ = \;& K^2 + \left[\int k^2 dF_0(k) - K^2\right] \exp\left\{-\int_0^t m(\varepsilon_z, \varepsilon_z) \frac{H_z(H_z + 2\kappa\varepsilon_z)}{2(H_z + \kappa\varepsilon_z)^2} dz - \frac{\alpha\nu}{2} t\right\} \end{aligned}$$

Thus the solution to $J_0(\alpha)$ can be written as

$$J_0(\nu) = \frac{\alpha}{4}\left[\int k^2 dF_0(k) - K^2\right] \int_0^T e^{-rt} H_t \exp\left\{-\int_0^t m(\varepsilon_z, \varepsilon_z) \frac{H_z(H_z + 2\kappa\varepsilon_z)}{2(H_z + \kappa\varepsilon_z)^2} dz - \frac{\alpha\nu}{2} t\right\} dt.$$

Thus the equilibrium matchmaking is

$$\psi = \frac{\alpha}{4}\left[\int k^2 dF_0(k) - K^2\right] \int_0^T e^{-rt} H_t \exp\left\{-\int_0^t m(\varepsilon_z, \varepsilon_z) \frac{H_z(H_z + 2\kappa\varepsilon_z)}{2(H_z + \kappa\varepsilon_z)^2} dz - \frac{\alpha\nu}{2} t\right\} dt. \tag{2.88}$$

The following proposition characterizes the comparative statics of the equilibrium measure of brokers with respect to policy and technology parameters.

Proposition 2.13 *Suppose κ is sufficiently small. The comparative statics of ν are*

	i^{ER}	i^{DW}	K
ν	$-$	$+$	$-$

Proof. Note that $J_0(\infty) = 0$ and $J_0(0) > 0$. For the existence of equilibrium we assume $\psi < J_0(0)$. Due to free entry, we focus on the equilibrium ν^* with $J_0'(\nu^*) < 0$. We define

$$M_t \equiv e^{-rt} H_t \exp\left\{-\int_0^t m(\varepsilon_z, \varepsilon_z) \frac{H_z(H_z + 2\kappa\varepsilon_z)}{2(H_z + \kappa\varepsilon_z)^2} dz - \frac{\alpha\nu}{2} t\right\},$$

which implies

$$\frac{\dot{M}_t}{M_t} = -\frac{3}{8}\alpha\nu - \frac{a_2}{H_t} - \frac{m\left(\varepsilon_t, \varepsilon_t\right)}{4}\frac{H_t}{H_t + \kappa\varepsilon_t}\left(\frac{2\kappa\varepsilon_t}{H_t + \kappa\varepsilon_t} + 1\right) < 0,$$

and

$$\frac{\partial}{\partial H_t}\left(\frac{\dot{M}_t}{M_t}\right) = \frac{a_2}{H_t} - \frac{m\left(\varepsilon_t, \varepsilon_t\right)}{4}\frac{\kappa\varepsilon_t}{\left(H_t + \kappa\varepsilon_t\right)^2}\left(\frac{4\kappa\varepsilon_t}{H_t + \kappa\varepsilon_t} - 1\right).$$

Thus a sufficient condition for $\frac{\partial}{\partial H_t}\left(\frac{\dot{M}_t}{M_t}\right) > 0$ is $\frac{\kappa\varepsilon_t}{H_t + \kappa\varepsilon_t} < \frac{1}{4}$, which requires a sufficiently small κ. Given this condition and note that $M_0 = H_0$, we can obtain that the path of M_t shifts upward if the path of H_t shifts upward. Combining with the result that $\frac{\partial H_t}{\partial A_2} > 0$, we can obtain that

$$\frac{\partial J_0\left(\nu\right)}{\partial A_2} = \int_0^T \frac{\partial J_0\left(\nu\right)}{\partial M_t}\frac{\partial M_t}{\partial A_2}dt > 0.$$

By implicit function theorem, we can obtain $\frac{\partial \nu^*}{\partial A_2} > 0$. Given $\frac{\partial A_2}{\partial i^{ER}} < 0$, $\frac{\partial A_2}{\partial i^{DW}} > 0$ and $\frac{\partial A_2}{\partial K} < 0$, this establishes our proposition. **Q.E.D.**

2.F Appendix: Payment shocks

Since Poole (1968) there has been a long history of analyzing the effects of payment flow on the Federal funds market. In this extension we study the role of payment on disintermediation. Suppose that banks are receiving and sending exogenous and stochastic payment flows of reserve balances. There are two types of payment flows: lumpy or continuous. Lumpy payments occur occasionally at the arrival rate ζ, with the amount w (negative value means outflow of reserve balances) drawn from a symmetric distribution G with mean 0 and standard deviation σ_L. Continuous payments occur continuously that follows a Brownian motion with mean μ and volatility σ_C. Thus the aggregate inflow of reserve balances from payment

flow is μ. The HJB equation becomes

$$\begin{aligned} rV_t(k) &= \dot{V}_t(k) + u(k) + \max_{\varepsilon\in[0,1]} \int \frac{1}{2} S_t(k,k',\varepsilon,\varepsilon_t(k'))\, m(\varepsilon,\varepsilon_t(k'))\, dF_t(k') \\ &\quad +\zeta \int [V_t(k+w) - V_t(k)]\, dG(w) + \mu \frac{\partial}{\partial k} V_t(k) + \frac{\sigma_C^2}{2} \frac{\partial^2}{\partial k^2} V_t(k). \end{aligned} \tag{2.89}$$

Given $\{F_t\}$, the value function in an equilibrium is given by

$$V_t(k) = -H_t k^2 + E_t k + D_t, \tag{2.90}$$

where H_t, E_t and D_t are given by the solutions to the following initial-value ODE problems

$$\dot{H}_t = rH_t - a_2 + \frac{1}{4} \frac{H_t^2 [(\lambda - \lambda_0)\varepsilon_t^2 + \lambda_0]}{\kappa\varepsilon_t + H_t}, \tag{2.91}$$

$$\dot{E}_t = rE_t - a_1 + \frac{K_t}{2} \frac{H_t^2 [(\lambda - \lambda_0)\varepsilon_t^2 + \lambda_0]}{\kappa\varepsilon_t + H_t} + 2\mu H_t, \tag{2.92}$$

$$\dot{D}_t = rD_t - \frac{1}{4} \frac{H_t^2 [(\lambda - \lambda_0)\varepsilon_t^2 + \lambda_0]}{\kappa\varepsilon_t + H_t} \int k'^2 dF_t(k') + \left(\zeta\sigma_L^2 + \sigma_C^2\right) H_t - \mu E_t, \tag{2.93}$$

where $H_T = A_2$, $E_T = A_1$ and $D_T = 0$. The equilibrium search profile of $\Omega(S_t, F_t)$ is given by

$$\varepsilon_t(k) = \begin{cases} 1, & \text{if } H_t \geq \eta; \\ 0, & \text{otherwise.} \end{cases} \tag{2.94}$$

The Federal funds purchased $q_t(k,k')$ and the Federal funds rate $\rho_t(k,k')$ are given by

$$q_t(k,k') = \frac{H_t(k'-k)}{2(\kappa\varepsilon_t + H_t)}, \tag{2.95}$$

$$\rho_t(k,k') = e^{r(T+\Delta-t)} [E_t - H_t(k+k')]. \tag{2.96}$$

Note that H_t does not depend on the payment shocks, while E_t is only affected by μ. The following Proposition summarize the grid-locking effect of payment shocks.

Proposition 2.14 *The comparative statics of the length of search, $\bar{\tau}$, the amount of Federal funds purchased, $q_t(k,k')$ and the Federal fund rates, $\rho_t(k,k')$ are given by the following table*

	$\bar{\tau}$	$q(k,k',\tau)$	$\rho(k,k',\tau)$
ζ	0	0	0
μ	0	0	$-$
σ	0	0	0

Proof. Since H_t is independent of the payment shocks, the comparative statics of $\bar{\tau}$ and q_t over payment shock parameters are zero. For ρ_t, the comparative statics is non-zero only for μ. Note that a higher μ means a higher K_t and a larger $2\mu H_t$. This implies a larger $\dot{E}_t$. Since E_T is given, it means E_t decreases in μ. Thus ρ_t decreases in μ. **Q.E.D.**

Intuitively, a larger μ means the excess reserves increase faster. This implies a lower marginal value of holding reserves, leading to lower Federal funds rates.

2.G Appendix: Counterparty risk

Afonso et al. (2011) documents the importance of counterparty risk in explaining the rise of Federal funds rate and decline Federal funds trade during the crisis. Our model can be extended to incorporate two kinds of counterparty risk. Consider that there is probability $1-p_L$ that, after the terms of trade is determined, the Federal funds lender cannot deliver the corresponding reserves to the borrower and the trade has to be cancelled. Also, there is a probability $1-p_B$ that the Federal funds borrower cannot repay R when it is due. The borrower's surplus is thus given by

$$p_L\left[V_t(k+q)-p_B e^{-r(T+\Delta-t)}R\right]-p_L V_t(k)-\chi(\varepsilon,q).$$

The lender's surplus is given by

$$p_L\left[V_t\left(k'-q\right)+p_B e^{-r(T+\Delta-t)}R\right]-p_L V_t\left(k'\right)-\chi\left(\varepsilon',-q\right).$$

The solution to Nash bargaining problem becomes

$$q_t\left(k,k',\varepsilon,\varepsilon'\right)=\frac{H_t\left(k'-k\right)}{\kappa/p_L\cdot\left(\varepsilon+\varepsilon'\right)+2H_t}, \tag{2.97}$$

$$R_t\left(k,k',\varepsilon,\varepsilon'\right)=\frac{e^{r(T+\Delta-t)}}{p_B}\left[E_t-H_t\left(k+k'\right)-\frac{\kappa\left(\varepsilon-\varepsilon'\right)}{2p_L}q_t\left(k,k',\varepsilon,\varepsilon'\right)\right]q_t\left(k,k',\varepsilon,\varepsilon'\right), \tag{2.98}$$

$$\rho\left(k,k',\tau\right)=\frac{R\left(k,k',\tau\right)}{q\left(k,k,\tau\right)}=\frac{e^{r(T+\Delta-t)}}{p_B}\left[E_t-H_t\left(k+k'\right)-\frac{\kappa\left(\varepsilon-\varepsilon'\right)}{2p_L}q_t\left(k,k',\varepsilon,\varepsilon'\right)\right], \tag{2.99}$$

$$S_t\left(k,k',\varepsilon,\varepsilon'\right)=\frac{p_L\left[H_t\left(k'-k\right)\right]^2}{\kappa/p_L\cdot\left(\varepsilon+\varepsilon'\right)+2H_t}. \tag{2.100}$$

The optimal search intensity in the most liquid equilibrium is

$$\varepsilon_t=\begin{cases}1, & \text{if } H_t\geq\frac{\kappa}{p_L}\left[\frac{\lambda}{2(\lambda-\lambda_0)}-1\right];\\ 0, & \text{otherwise.}\end{cases}$$

The solution to the value function is that $V_t\left(k\right)=-H_t k^2+E_t k+D_t$, where

$$\dot{H}_t = rH_t-a_2+\frac{p_L}{4}\frac{H_t^2}{\kappa/p_L\cdot\varepsilon_t+H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2+\lambda_0\right],\ \text{where } H_T=A_2; \tag{2.101}$$

$$\dot{E}_t = rE_t-a_1+\frac{Kp_L}{2}\frac{H_t^2}{\kappa/p_L\cdot\varepsilon_t+H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2+\lambda_0\right],\ \text{where } E_T=A_1; \tag{2.102}$$

$$\dot{D}_t = rD_t-\frac{p_L}{4}\frac{H_t^2}{\kappa/p_L\cdot\varepsilon_t+H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2+\lambda_0\right]\int k'^2 dF_t\left(k'\right),\ \text{where } D_T=0 \tag{2.103}$$

Overall, the effects of higher counterparty risk (a higer $1-p_L$) are isomorphic to the effects of higher transaction cost κ and lower matching rate λ and λ_0.

2.H Appendix: Algorithm of simulation and estimation

Simulation. Let us denote N as the number of banks, $i \in \{1, 2, ..., N\}$ as the index of individual banks. Since the size of peripheral traders is redundant for simulation, we assume there is only one peripheral trader and denote it as $i = N + 1$. We also denote $m \in \mathbb{N}$ as the index for bilateral meetings, where a smaller m means an earlier meeting. Since the number of banks is finite, the total number of meetings is also finite. Moreover, denote $k_0(i)$ as the initial reserve balances of bank i before entering the Federal funds market, and $k_m(i)$ as the reserve balances of bank i after meeting m takes place. Note that $k_0(i)$ is given by banks' empirical excess reserves divided by bank assets, and $k_m(i) \neq k_{m-1}(i)$ only if bank i is one of the counterparties in meeting m. It is important to note that the mass of an individual bank is normalized to 1, and the search intensity λ, λ_0 and φ represent the search intensity for an individual bank. Thus the total mass of banks is N, and the contact rate for a bank with another bank is $\frac{m(\varepsilon_t, \varepsilon_t)}{N}$. There are $\frac{N(N-1)}{2}$ pairs of bilateral meetings between banks and N pairs of bilateral meetings between a bank and a peripheral trader. All these meetings are independent Poisson process. Thus the sum of all these meetings follows a Poisson process with intensity $\frac{N(N-1)}{2}\frac{m(\varepsilon_t, \varepsilon_t)}{N} + N\varphi = \frac{N-1}{2} m(\varepsilon_t, \varepsilon_t) + N\varphi$. We simulate the discretized version of the model via the following algorithm.

1. Given the model parameters and policy parameters, we numerically solve the paths of H_t, ε_t via Proposition 2.12 and solve the paths of E_t and K_t via the ODEs (2.83) and (2.84).

2. Given the path of ε_t, simulate a Poisson process for bilateral meetings up to time T via a thinning algorithm:[23]

   (a) Set a sufficiently large $\lambda_{\max}$ (such that $\lambda_{\max} > \lambda$). Generate a random integer $\hat{M}$ distributed as Poisson with mean $\left(\frac{N-1}{2}\lambda_{\max} + N\varphi\right) T$. If $\hat{M} = 0$ stop.

[23] See Sigman (2007) for a detailed description and proof of the thinning algorithm.

(b) Generate $\hat{M}$ random numbers distributed as i.i.d. uniforms on $(0,1)$, i.e. $U_1, ..., U_{\hat{M}}$, and reset $U_m = T \cdot U_m$, $m \in \left\{1, ..., \hat{M}\right\}$.

(c) Place the U_m in ascending order to obtain the order statistics $U_{(1)} < U_{(2)} < ... < U_{(\hat{M})}$.

(d) Set $\hat{t}_m = U_{(m)}$.

(e) For each $\hat{t}_m$, generate an i.i.d. uniform on $(0,1)$, $\hat{U}_m$. If

$$\hat{U}_m \leq \frac{\frac{N-1}{2} m\left(\varepsilon_{\hat{t}_m}, \varepsilon_{\hat{t}_m}\right) + N\varphi}{\frac{N-1}{2}\lambda_{\max} + N\varphi},$$

then keep $\hat{t}_m$. Otherwise, drop it.

(f) For each kept $\hat{t}_m$, draw a pair of integers $\hat{p}_m = \{i, j\}$ with $1 \leq i < j \leq N+1$ from the weighted distribution j

$$\Pr(i,j) = \begin{cases} \frac{\frac{N-1}{2} m\left(\varepsilon_{\hat{t}_m}, \varepsilon_{\hat{t}_m}\right)}{\frac{N-1}{2} m\left(\varepsilon_{\hat{t}_m}, \varepsilon_{\hat{t}_m}\right) + N\varphi} \frac{2}{N(N-1)}, & \text{if } i, j \leq N, \\ \frac{N\varphi}{\frac{N-1}{2} m\left(\varepsilon_{\hat{t}_m}, \varepsilon_{\hat{t}_m}\right) + N\varphi} \frac{1}{N}, & \text{if } j = N+1. \end{cases}$$

(g) For each kept $\hat{t}_m$ and $\hat{p}_m$, we relabel them with $\{t_m, p_m\}_{m=1}^{M}$, where $t_m < t_{m+1}$ and M is the number of kept $\hat{t}_m$. The sequence of $\{t_m, p_m\}_{m=1}^{M}$ is the Poisson process for bilateral meetings for our simulation. and denote the number of kept trade according to the rule. Update $k_n(i)$ and $k_n(j)$.

3. Update individual reserve balances and bilateral terms of trade: denote $k_m(i)$, $q_m(i)$ and ρ_m as bank i's reserve balances after meeting m, bank i's cumulative absolute Federal funds trade after meeting m, and the bilateral Federal funds rate in meeting m. We start with the data $k_0(i)$ and set $q_0(i) = 0$ by definition. For each meeting m, if $i \in p_m$, then update $k_m(i)$, $q_m(i)$ and ρ_m according to the theoretical formulae. For any $i \notin p_m$, do not update $k_m(i)$ and $q_m(i)$.

4. Use the sequence $\{k_m(i), q_m(i), \rho_m\}$ to calculate the aggregate moments and regression coefficients.

Estimation. The simulated method of moments estimation follows a standard two-step procedure.[24] For each quarter, we simulate the model for $S = 2,000$ times.

[24]See Adda and Cooper (2003) for the reference on simulated method of moments.

CHAPTER 3

Transfers vs Credit Policy: Macroeconomic Policy Trade-offs during Covid-19

with Saki Bigio and Eduardo Zilberman

3.1 Introduction

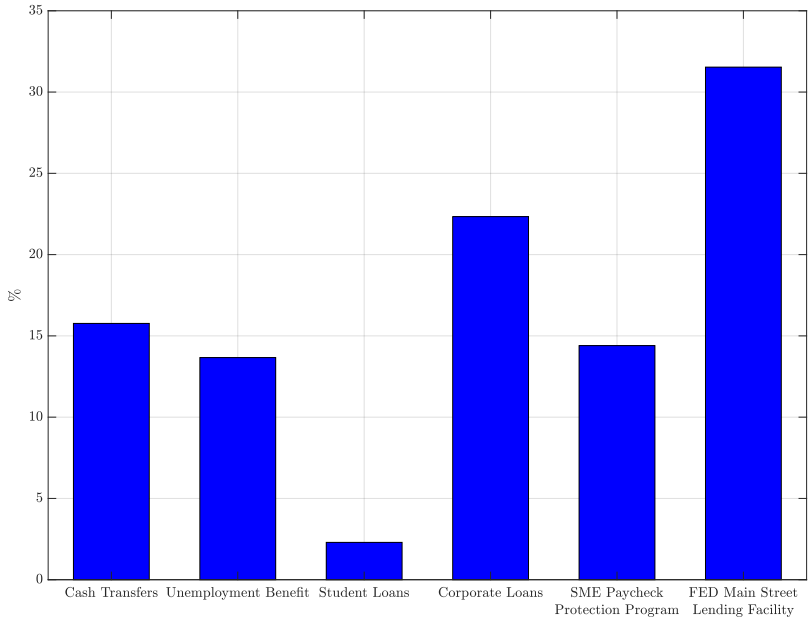
The Covid-19 pandemic is a quintessential macroeconomic shock. Large and unexpected, the shock has escalated so quickly that the self-stabilizing mechanisms of business cycles are likely not to work. Learning from the recent experiences of the Great Recession, central banks and fiscal authorities have responded with unprecedented speed and scale. An equally unprecedented amount of policy recommendations has been produced by the macroeconomics community, e.g. Brunnermeier et al. (2020), Gourinchas (2020), among many others. Calls for macroeconomic-stabilization have primarily focused on large scale transfer programs and central-bank open-market operations that facilitate bank credit. Both programs are geared toward expanding the amount of social insurance, while acting as a demand stabilizer.

In the case of the US, the combination of the CARES act setup by the US Treasury and the Main Street Lending Facility of the Federal Reserve offer amount to a combination of 1,902 billion US dollars in direct transfers, unemployment insurance, and credit to firms. A decomposition of this amount in this subset of programs is presented in Figure 3.1.

Internationally, a recent report by UBS[1] claims that 3.7 percent, so far, of global GDP has

[1] "Global Economic Perspectives. Bubble, Bubble, Toil, and Trouble: which fiscal mix will work against Covid-19?," UBS Global Research Team, 21 April 2020.

Figure 3.1: US Covid-19 Policy Response: Decomposition of Programs Directed to Consumers and Loans



Notes: This figure reports the share of each component of a subset of the CARES act programs and the Main Street Lending Facility of the Federal Reserve. All figures are presented in current US Dollars: The Fed's Main Street Lending Facility amounts to 600 Billion. The CARES act assigns Cash Transfers for 300 Billion, 260 Billion in Unemployment Benefits and 43.7 Billions in Student Loans (as part of the Economic Impact Payments to American Households program), 274 Billion in loans to small business (as part of the Small Business Paycheck Protection Program) and 425 Billion in Corporate Loans. The figures are interpreted from statements from the U.S. Department of the Treasury: `https://home.treasury.gov/policy-issues/cares/preserving-jobs-for-american-industry`, and the classifications cross-checked with related articles: `https://messertodd.github.io/coronavirus-policy-response/Unconventional-MP.html`, and `https://www.npr.org/2020/03/26/821457551/whats-inside-the-senate-s-2-trillion-coronavirus-aid-package`.

been put forward in policies that vary in scale and target across countries. There is a striking difference between the response of developed economies, and other economies. Developed markets are usings 31 percent of fiscal expenditures on job retention schemes, 15 percent on business loans and grants, 15 percent on tax reliefs, 12 percent on direct cash payments, mostly driven by Japan where the share of direct transfers is 57 percent, and 10 percent on

unemployment insurance. In emerging economies without China,[2] the figures on tax reliefs and unemployment insurance are comparable to those in developed economies, 18 percent and 10 percent, respectively. Nonetheless, emerging markets are relying more on direct cash payments (20 percent), but less on job retention schemes (13 percent) and business loans and grants (4 percent). These figures are subject to revisions and reclassification, and may change as the crisis evolves.

As policies are being refined on the spot, there is room for simple theories that assist this analysis. Here is one such environment. The paper presents a simple incomplete markets economy, which we use to frame the recent discussions regarding the macroeconomic trade-offs related to the policy responses during the Covid-19 crisis. The model is an extension of Bigio and Sannikov (2021), with two sectors. The economic forces are similar in spirit to Guerrieri et al. (2020). One sector produces social goods, which need social interactions to be consumed, whereas the other produces goods that can be consumed remotely. Households are subject to an unexpected shock that reduces the utility from consuming social goods. After the shock is announced, except for the idiosyncratic risk, the economy proceeds in a deterministic fashion. This reduction in utility is due to the fear of being infected, which induces a behavioral response of individuals during a pandemic.

The Covid-19 manifests as a discount factor shock in an equivalent representation of the model with one sector. If the elasticity of substitution between social and remote goods (ESBG) as well as the elasticity of intertemporal substitution (EIS) are greater than one (something we assume throughout the paper), the shock can be represented equivalently as a positive discount factor shock (less discounting). Since we assume labor is supplied inelastically and there is perfect reallocation of labor input across sectors, under the flexible prices benchmark, after the shock the remote good sector absorbs the fall in the social good sector. The unemployment rate remains at its natural level. Once we introduce rigid wages,

[2]China is an outlier as its response to the crisis concerns mostly public investment, which according to the report represents 59 percent of total expenditures.

if EIS is sufficiently high (or ESBG sufficiently small), a point made in Guerrieri et al. (2020), the unemployment rate adjusts to accommodate the decline in aggregate demand that results from the shock. This decline in consumption may manifest in both sectors, opening the door for macroeconomic stabilization. Inspired by the current policy debate, we study the effectiveness of lump-sum transfers versus a credit policy that generates the same path of government indebtedness.

We argue that the power of lump-sum transfers versus a credit subsidy policy critically depends on the level of financial development. Take two extremes. If we consider a natural borrowing limit, the Ricardian equivalence holds and lump-sump transfers are neutral. In this case, credit policy should be the preferred tool, and we show that it mitigates the recession. Now consider an economy with the borrow limit equal to zero. In this case, a credit policy is immaterial and lump-sum transfers are the preferred instrument. We also showcase an intermediate case. The choice of whether to go for transfers or credit policy (or the optimal mix between them) depends crucially on how the lending channel is impaired due to Covid-19 crisis, captured here by some static comparative on the borrowing limit. A first lesson is that economies with a developed financial system should unleash credit. Developing economies should rely more on transfers. We also argue that a stringent debt limit amplifies the recession, at the same time that restricts the use of credit subsidy, perhaps the preferred instrument to target those households who really need support. We enrich this discussion presenting a numerical illustration where we compare both policies, for different levels of debt limits. To make policies comparable, we assume they produce the same path of real government liabilities.

This is a rudimentary environment that, being so, faces several limitations. An obvious limitation is the assumption of perfect labor reallocation across sectors, given that one feature of the crisis is that the actual mass of unemployed cannot be easily absorbed by the remote sector. The model assumes that a bar tender can work in grocery deliveries the next day. Yet, the lesson is valid: the goal of macroeconomic stabilization is to prevent the needed recession

in social sectors to spill over to sectors that can be maintained active. Second, the policies themselves are very stark. For example, transfers are not directed to the poor, which are likely to contract their demand the most for precautionary reasons. Targeted transfers are part of social insurance in many countries and their effects in incomplete market economies are well understood (e.g. Berriel and Zilberman, 2011). However, the speed of the crisis limits the ability to setup transfers targeted to low wealth households rather than low income households. Thus, transfers can be targeted only in as much income is a proxy for wealth. The credit policy here is also stark because it targets individual debt only. In practice, governments are also attempting to target firm credit, as a means to keep workers in their jobs and avoid the social losses produced by inefficient unemployment spells. We explain how the model can be modified easily along those lines. Yet, we think that governments should attempt to stimulate credit card debt, a policy that has not being used as much as others. Third, the debt limits are exogenous. This is a key feature of the model that determines the relevance of the transfers versus credit subsidy, but clearly a credit program has to confront the reality of default risk and moral hazard. Finally, we leave aside the question of how these programs are financed. Temporary deviations from a Taylor rule allow, at least partially, the consolidated government to control the trajectory of real rates, and this can potentially introduce trade-offs.

A final important shortcoming is that the model is not tailored to speak to forced lockdowns (or any other containment policy). This could be simply accommodated by an ad-hoc restriction on the amount of social goods that can be consumed. The representation in terms of the discount factor would be the same as the one here, but the representation would be endogenous to policy. These are all issues that can be addressed in variations to this model. Despite these limitations, and the distance to reality, the model here is a good laboratory to think of policy implications in real time.

This paper is one of the many research responses that study the positive and normative macroeconomic implications of the Covid-19 pandemic. One group of papers integrates

epidemiology and macro models,[3] and study how the evolution of the pandemic interacts with the macroeconomy. Examples include Eichenbaum et al. (2020), Alvarez et al. (forthcoming), Jones et al. (forthcoming), Bethune and Korinek (2020), Krueger et al. (2020) and Kaplan et al. (2020). Another group interprets the Covid-19 shock by using the standard macroeconomics toolkit. In particular, the shock is interpreted as an unexpected shut down of part of the economy as a consequence of the pandemic shock and unmodeled sanitation measures. In Guerrieri et al. (2020), the shock is represented as a cap on labor employment. A key finding is that this supply shock only generates demand deficiency in a tow-sectors economy for a specific (but not restrictive) set of parameters. In Caballero and Simsek (forthcoming), the large supply shock triggers a stock wealth decline of risk-tolerant agents, leading to a feedback loop that further decreases asset prices and output when the interest rate is constrained downward to accommodate the shock. In Buera et al. (2021), some of the firms in their model with heterogeneous entrepreneurs must shut down due to the shock. Our paper fits this branch of this emerging literature. We represent the Covid-19 shock as a shock that reduces the consumption of social goods. We interpret it as a decease in marginal utility of consuming such goods due to behavioral responses to the risk of being infected. Incidentally, this shock is more likely to be recessive for a specific set of parameters, similar to the set found by Guerrieri et al. (2020). Closely related is Faria-e Castro (2021), who studies policy responses to a similar shock in a two-sector model with two types of agents, borrowers and lenders. He finds that unemployment insurance is the preferred policy from the point of view of borrowers, while savers favor transfers. Finally, there is a third branch that studies empirically the macroeconomic consequences of early pandemics. For example, Barro et al. (2020), Correia et al. (2020), and Jordà et al. (2020). This literature is evolving fast, and these papers represent a rather incomplete list.

The paper is organized as follows. Section 2 presents the model. Section 3 discusses the

[3]A description of the baseline epidemiology model and simulations concerning the Covid-19 pandemic can be found in Atkeson (2020) and Berger et al. (2020).

results. Finally, Section 4 concludes.

3.2 Model

Time is continuous, $t \in [0, \infty)$. The economy is populated by ex-post heterogeneous households. There is a consolidated government, which is a combination of a Central Bank (CB) and fiscal authority. Banks intermediate between borrower and lender households, but since they make zero profits, they are simple pass through entities. The CB determines a common policy rate and conducts open market operations that translate into a credit subsidy. The fiscal authority makes/collects (lump sum) transfers/taxes to/from households and manages unemployment insurance. Households face idiosyncratic income shocks produced by unemployment spells. Households self-insure by borrowing and lending through banks.

There are two goods produced with a common input, labor. Due to wage rigidities, a shock aimed to capture the economic impact of the Covid-19 pandemics (i.e., a shock that reduces the consumption of goods that require social contact) generates unemployment beyond a natural level. This social inefficiency is amplified through debt constraints. The objective is to study the role of two specific policies aimed at stabilizing the economy after the shock. The policy options are either a credit policy that shows up as a lending subsidy or lump-sum direct transfers to households.

3.2.1 Preferences, technology, and the Covid-19 shock

To accommodate a Covid-19 shock, we consider two types of goods, one that can be consumed remotely, c_t^r, and another that requires social interactions, c_t^s. Let households instantaneous preferences be given by $U(x_t) \equiv \left(x_t^{1-\gamma} - 1\right) / (1-\gamma)$; the composite good x_t is given by

$$x_t = \left(\alpha^{1/\epsilon} c_t^{r\,1-1/\epsilon} + \left((1-\alpha)\,\beta_t\right)^{1/\epsilon} c_t^{s\,1-1/\epsilon}\right)^{\epsilon/(\epsilon-1)}.$$

Here, γ is both the risk aversion parameter and the inverse of the intertemporal elasticity of substitution. For this version, the risk-aversion force is dominated by the elasticity of substitution. In turn, ϵ is the elasticity of substitution between goods, and α is the share of each type of good. Throughout the paper, we take a stance that social and remote goods are substitute, $\epsilon > 1$, such that the consumption of remote goods (at home, e.g., Netflix, Amazon, supermarket expenditures, etc) increases, whenever the consumption of social goods (e.g., movies, restaurants, etc) decreases.

We interpret $\beta_t \in [0, 1]$ as a shock that reduces the utility from social goods due to the fear of the pandemic. Intuitively, individuals arguably experiment a lower degree of happiness whenever consuming certain goods that involve risk of being infected. In that sense, our model is capturing a behavioral response of individuals during a pandemic, rather than a proper containment policy that imposes social distance. Alternatively, we could impose an upper limit on the consumption of social goods, say $c_t^s \leq \overline{c}^s$, that would capture a policy that aims to limit social interactions, such as a lockdown. Both assumptions show up mathematically as a modification to the expenditure problem, but in the case of rationing, that representation would not be invariant to policy. We will study a quantity restriction in a new version of the paper. Utility flows are discounted accordingly with discount rate ρ, $\mathbb{E}\left[\int_0^\infty e^{-\rho t} U\left(x_t\right) dt\right]$.

Let c_t be total expenditures. For simplicity, we assume that both goods can be produced with the same production function, so their relative prices is one—we can relax this assumption without difficulty. The first-order conditions and simple algebra yield:

$$c_t^r = \frac{\alpha}{\left(\left(1-\alpha\right)\beta_t\right)+\alpha} c_t; c_t^s = \frac{\left(\left(1-\alpha\right)\beta_t\right)}{\left(\left(1-\alpha\right)\beta_t\right)+\alpha} c_t; x_t = \left(\left(1-\alpha\right)\beta_t + \alpha\right)^{1/(\epsilon-1)} c_t.$$

Naturally, a decrease in β_t not only decreases (increases) the consumption of social (remote) goods, but if the elasticity of substitution is more than one, $\epsilon > 1$, for a given amount of expenditures, the consumption of the composite good decreases. For plausible assumptions

on α and ϵ, conceptually we can reverse engineer β_t to deliver the path of c_t^s compatible with people fearing going out and staying at home for a given time, or the path desired by a containment policy $\overline{c}^s$ that limits consumption of social goods $c_t^s = \overline{c}^s$.

By substituting $x_t = \left((1-\alpha)\,\beta_t + \alpha\right)^{1/(\epsilon-1)} c_t$ back into the households' objective function, and given that preferences are CRRA, one obtains

$$\mathbb{E}\left[\int_0^\infty e^{-\rho t}\xi_t U\left(c_t\right) dt\right],$$

where $\xi_t = \left((1-\alpha)\,\beta_t + \alpha\right)^{(1-\gamma)/(\epsilon-1)}$.

For $(1-\gamma)/(\epsilon-1) < 0$, which is true, for instance, when both $\epsilon > 1$ and $\gamma > 1$, a Covid-19 shock that reduces the consumption of social goods, morphs into a negative discount factor shock (more discounting of the future). In the other case, in which $\epsilon > 1$ and $\gamma < 1$, the shock can be represented equivalently as a positive discount factor shock (less discounting) as agents do not gain much utility from current consumption. A positive discount factor shock, coupled with price rigidity, tends to generate a recession whenever the intertemporal elasticity of substitution is more than one, $IES = 1/\gamma > 1$, as agents are willing to substitute aggregate consumption intertemporally. Hence, we also assume throughout the paper that $\gamma < 1$, so our exercise can easily reproduce the feature that the Covid-19 shock is clearly recessive. Note that the assumption that $\epsilon > 1 > \gamma$ is consistent with the set of parameters for which Guerrieri et al. (2020) find that a supply shock in one of the sectors in their two-sectors model generates demand deficiency, something those authors calls a Keynesian supply shock—here we simply refer to the spill-over across sectors. Finally, note that the higher the share of social goods, $(1-\alpha)$, the more intense will be the propagation of the shock.

In addition, and for simplicity, we assume that production is a linear function of aggregate employment, $c_t^r + c_t^s = (1 - U_t)$, where U_t is the unemployment rate. Note that labor can be perfectly reallocated across sectors in response to the Covid-19 shock. This is a useful

benchmark to show that, even under this assumption, the shock can induce a recession with a scope for policy response aimed at stabilizing output. Nonetheless, imperfect reallocation seems to be a key ingredient under the current crisis: given that the mass of unemployed cannot be easily absorbed by the essential sector, which would further increase the scope for policy to improve welfare.

Finally, there is no notion of payroll financing in this paper, as firms do not face frictions and always honor their wage bills. Although subsidizing firms is an important aspect of the current crisis, the focus here is on the trade-off between lump-sum transfers and credit subsidy to households, rather than firms. Nonetheless, we can easily modify this approach, by studying a subsidy policy to keep workers away from unemployment, but keeping them in a different pool of idle firms. The idea is that these workers could return to the workforce immediately, rather than going through the unemployment to employment process. We think this is a core policy response, but for now we abstract from it.

3.2.2 Households

The non-financial sector features a measure-one continuum of ex-ante identical households, that are ex-post heterogeneous. Heterogeneity follows from uninsurable idiosyncratic risk $z \in \{u, e\}$, where u stands for unemployed and e stands for employed. The stochastic process that governs this idiosyncratic risk is independent and identically distributed across households. Households transit from one state to another according to an instantaneous transition probabilities of $\Gamma_t^{eu} = \nu^{eu} + \phi_t^+$ and $\Gamma_t^{ue} = \nu^{ue} - \phi_t^-$, where $\{\nu^{ue}, \nu^{eu}\}$ are exogenous (or natural) transition rates, and ϕ_t is an endogenous employment-unemployment adjustment rate that occurs due to price rigidity. Namely, ϕ_t is positive (negative) when the rigidity constraint is binding and there is an excess supply (demand) of final goods under $\phi_t = 0$.

Households receive a flow of real income given by:

$$dw_t = y(z)\,dt + T_t dt,$$

where income w_t is the sum of direct transfers T_t (to be described below) and labor income (or unemployment insurance) $y(z)$. Note that, in equilibrium, the real wage rate is one. We assume that $y(e) = \left(1 - \tau^l\right)$, and thus, households are taxed with rate τ^l whenever they are employed, and $y(u) = b$, meaning that b is the replacement rate for unemployment insurance purposes.

Financial claims are nominal. Although all claims are nominal, the individual state variable is s_t, the stock of real financial claims—the distinction would matter only with long-term debt. Households store wealth in bank deposits, a_t^h, or currency, m_t^h, and borrow loans against banks, l_t^h. By convention, $\left\{a_t^h, m_t^h, l_t^h\right\} \geq 0$. Let real rates of return be $r_t^m \equiv i_t^m - \pi_t$ and $r_t^l \equiv i_t^l - \pi_t$, where i_t^m is the monetary policy rate, i_t^l is the rate on loans, and $\pi_t = \dot{P}_t / P_t$ is inflation. Note that the price of the composite good in terms of money is P_t. Currency does not yield nominal interest, and thus, its real return is $-\pi_t$. The law of motion for real wealth follows

$$ds_t = \left(r_t^m \frac{a_t^h}{P_t} - \pi_t \frac{m_t^h}{P_t} - r_t^l \frac{l_t^h}{P_t} - c_t \right) dt + dw_t, \tag{3.1}$$

and the balance-sheet identity is given by

$$\frac{a_t^h + m_t^h}{P_t} = s_t + \frac{l_t^h}{P_t}.$$

From a household's perspective, there is no distinction between holding deposits or currency beyond their rates of return. Hence, currency is only held when the nominal deposit rate is zero, and both assets yield the same return. This feature is introduced into the model only to articulate a zero lower bound as an implementation constraint. Another observation is that households do not hold deposits and loans if there is a positive spread between them. Combining these insights, (3.1) can be written succinctly as:

$$ds_t = \left(r_t(s_t) s_t - c_t \right) dt + dw_t, \tag{3.2}$$

where $r_t(s_t) = r_t^m$ if $s_t \geq 0$, and $r_t(s_t) = r_t^l$ if $s_t < 0$. Finally, households can borrow, but only up to a certain limit. For now, there is a fixed threshold $\bar{s} \leq 0$, such that $s_t \geq \bar{s}$. In this version of the paper, we abstract from default risk and moral hazard associated with incentives to repudiate debt. This is arguably a relevant feature of the current crisis. As we discuss below, the degree of slackness of such debt limit is the key determinant of policy design in response to the shock. Hence, exogenous debt limit is another feature of the model that is clearly restrictive.

The Hamilton-Jabobi-Bellman (HJB) equation associated with the household's problem is:

Problem 3.1 *[Household's Problem]The household's value and policy functions are the solutions to:*

$$\begin{aligned} \rho V(z,s,t) &= \max_{\{c\}} \xi_t U(c) + V'(z,s,t)\left[r_t^m(s)s - c + y(z) + T_t\right] \\ &\quad + \Gamma_t^{zz'}\left[V(z',s,t) - V(z,s,t)\right] + \dot{V}(z,s,t), \end{aligned}$$

subject to $s \geq \bar{s}$.

3.2.3 Unemployment, inflation, and the Phillips curve

In standard Real Business Cycle models, prices adjust to ensure market clearing, whereas in standard New Keynesian (NK) models prices are rigid but firms produce and supply goods to meed their demand. Here, we follow the NK tradition, but instead of assuming monopolistic competition and a price adjustment process, we simply postulate that prices are rigid in a way that makes inflation evolve according to a classic forward-looking version of the Phillips curve:

$$\dot{\pi}_t = \rho\left(\pi_t - \pi_{ss}\right) - \kappa\left(U_{ss} - U_t\right),$$

where the subscript ss denote steady-state levels. In particular, inflation increases (decreases) whenever unemployment U_t is below (above) its natural rate U_{ss}, or inflation π_t is above (below) its long-run expected inflation-target π_{ss}, implemented by the CB interest-rate policy. Different from a NK model, steady-state inflation does not have consequences for efficiency.

By solving the equation forward, one obtains the following integral solution for inflation at time t,

$$\pi_t = \pi_{ss} + \kappa \int_0^\infty \exp(-\rho s)(U_{ss} - U_t)\, ds.$$

Importantly, π_t is not pre-determined, as it depends on path of future unemployment. Inflation is boosted at intensity κ, as unemployment falls below steady state. When unemployment is below steady state, the economy experience wage pressures. In that case, nominal wages tend to increase. Similarly, the economy features deflation as the unemployment rate rises above steady state.

In this model, as anticipated above, we assume that the endogenous component ϕ_t of employment-unemployment transitions adjusts to ensure market clearing. In particular, the law of motion of unemployment is given by

$$\dot{U}_t = \left[\nu^{eu} + \phi_t^+\right](1 - U_t) - \left[\nu^{ue} - \phi_t^-\right] U_t. \tag{3.3}$$

If demand is insufficient, and prices cannot adjust downwards, the unemployment rate responds.

3.2.4 Intermediation and implementation of credit spread with open-market operations

Financial intermediation is carried out by a competitive fringe of intermediaries. In the paper, these are simple pass through entities. In particular, banks choose their supply of

nominal deposits, a_t, nominal loans, l_t, and reserve holdings, m_t.[4] Free entry and perfect competition yield zero expected profits. Deposits and reserves earn corresponding rates $i_t^a = i_t^m$, where i_t^m is the policy target set by the CB. Due to the credit policy, the interest rate on loans, i_t^l, differs from the policy rate.

To implement the credit policy, we assume that the CB branch of the consolidated government conducts a policy that targets a given loan subsidy, $\sigma_t > 0$. This subsidy is induced by a combination of interest on reserves and open-market operations as explained below. Due to competition, i_t^l must be such that expected returns satisfy $i_t^l + \sigma_t = i_t^m$. In real terms, let interest rates be expressed as r_t^a, r_t^m and r_t^l. Given that $i_t^a = i_t^m = i_t^l + \sigma_t$, any composition of banks' balance sheet, such that $l_t + m_t = a_t$, is consistent with optimal behavior. The aggregate supply of deposits and loans, and holdings of reserves are denoted by A_t^b, L_t^b, and M_t^b, respectively, and of course, must satisfies $L_t^b + M_t^b = A_t^b$.

To explain how the CB implements a negative spread, we work with the continuous time limit of a discrete time implementation. To implement a spread σ_t, the CB purchases a fixed allotment of loans L_t^f. Loans are purchased at a random auction at a pre-specified price $q_t \equiv 1 + \Delta\sigma_t \cdot \frac{L_t^b}{L_t^f}$ for a small time interval Δ. Since the price is greater than one, all loans participate in the auction. Winners in the auction earn an arbitrage $(q_t - 1)$ per time interval. The probability of selling a loan is $\Theta = L_t^f / L_t^b \Delta$ per interval of time. Taking the interval to zero, then the bank earns an arbitrage of

$$\sigma_t \equiv \lim_{\Delta \to 0} (q_t - 1) \cdot \Theta_t,$$

per instant of time.

[4]Banks operate without equity. The introduction of a role for bank equity (via restrictions like capital requirements or limited participation) would produce bank profits and would make equity an aggregate state variable. For simplicity, we abstract away from this dimension in this paper. In practice, alleviating capital requirements is essential to expand credit.

3.2.5 Consolidated government

Throughout the paper, we consider the consolidated government budget constraint, which combines budget constraints associated with the Central Bank (CB) and the branch of the government responsible for fiscal policies, and use the terms CB and (consolidated) government interchangeably.

Budget constraint. The CB has a nominal net asset position, E_t, defined as:

$$E_t \equiv L_t^f - M_t.$$

The net-asset position is the difference between loans held by the CB, L_t^f, and the CB liabilities, i.e., the monetary base, M_t. The monetary base is divided into the aggregate holdings of reserves by banks M_t^b and the household's currency holdings, $M0_t$. The monetary base is always positive. The CB can issue or purchase loans L_t^f: when negative L_t^f is understood to be a stock of government bonds, when positive, it is understood to be the loan purchases of the CB.[5] An open market operation is a simultaneous increase or decrease in M_t and L_t^f without altering net asset positions.[6]

There are two "active" fiscal policies available to the government, lump-sum transfers T_t and the credit subsidy σ_t. The parameter b is important to control the degree of insurance in the economy and endow the unemployed with some income. We assume that the employment insurance is not necessarily balanced, leaving a deficit of $w_t b U_t - w_t \tau^l (1 - U_t)$. Considering

[5]There is no distinction between private and public loans. In fact, whenever $L_t^f < 0$, an increase in L_t^f is interpreted as a conventional open market operation (OMO). Instead, when $L_t^f > 0$, an increase L_t^f is an unconventional OMO. The assumption is that government bonds are as illiquid as private loans from the point of view of banks.

[6]The model is rich enough to accommodate a "helicopter drop" through an increase in M_t, without a counterpart increase in L_t^f.

the consolidated budget, the nominal fiscal surplus without transfers is given by:

$$\Pi_t^f = i_t^m L_t^f - i_t^m \left(M_t - M0_t\right) - P_t T_t - \sigma_t L_t^b + w_t \tau^l \left(1 - U_t\right) - w_t b U_t, \tag{3.4}$$

where the sources of income are given by the nominal interest on loans and labor taxes, whereas the sources of expenses are reserve remunerations, transfers, the credit subsidy and unemployment insurance. The net asset position of the consolidated government evolves according to

$$dE_t = d\Pi_t^f = \underbrace{dL_t^f - dM_t}_{\text{unbacked transfers}} .$$

The government accumulates a nominal claim on the private sector as undistributed income. The net asset position decreases with the difference between the monetary base and the loan purchases of the CB. In real terms, the CB's net asset position is $\mathcal{E}_t \equiv E_t/P_t$ and its loan holdings are $\mathcal{L}_t^f \equiv L_t^f/P_t$. Let $\mathcal{W}_t$ denote the real wage, and $f\left(z,s,t\right)$ the density associated with the joint distribution of savings s and employment status z. The next proposition exploits this observation to express the law of motion of the real net asset position $\mathcal{E}_t$ in real terms.

In real terms, $\mathcal{E}_t$ satisfies:

$$\dot{\mathcal{E}}_t = r_t^m \mathcal{E}_t + \left(\sigma_t - \pi_t\right) \int_{\bar{s}}^{0} s \left[f\left(e,s,t\right) + f\left(u,s,t\right)\right] ds + \mathcal{W}_t \left(\tau^l \left(1 - U_t\right) - b U_t\right) - \tau_t \mathcal{E}_t, \ \mathcal{E}_0 \text{ given.} \tag{3.5}$$

where transfers are given by $T_t = \tau_t \mathcal{E}_t$. The first term in (3.4) is the portfolio income earnings (losses) of the CB which equal the real rate times the net asset position. The second term captures the losses from the CB's subsidy. The third term is the outcome of income taxation and unemployment insurance policy. Finally, transfers are subtracted from the real asset position. A policy path is constrained by solvency conditions. An important restriction is a long-run solvency constraint for the CB. In particular, there is a limit $\lim_{t\to\infty} \mathcal{E}_t \geq \underline{\mathcal{E}}$ for some minimum $\underline{\mathcal{E}}$ that guarantees that the CB can raise enough revenues and satisfy

$d\underline{\mathcal{E}} = 0$. It must be the case that at $\underline{\mathcal{E}}$ discount window revenues cover any balance sheet costs. This condition is equivalent to assuming that the CB's liabilities are not worth zero in equilibrium. The model features a Laffer curve for CB revenues. Although we do not solve for $\underline{\mathcal{E}}$ explicitly, in the exercises we analyze in the following section, we impose that all policy paths lead to a convergent stable government net asset position and $\lim_{t\to\infty} d\mathcal{E}_t = 0$. Another restriction in the opposite direction is that $\mathcal{E}_t \leq -\bar{s}$, which is equivalent to saying that the CB claim on the public cannot exceed the public's debt limit.

Taylor rule. To set the interest **instensively** on reserves, the CB works with a Taylor rule that allows for a discretionary component that is triggered by the shock, but also follows a standard Taylor rule that captures commitment for the long-run. This feature is important. Without a Taylor rule, the model is unstable, so we need the long-run component. At the same time, we want to capture the idea that monetary policy responds to economic conditions. For that, we specify the following rule:

$$i_t^m = i_\infty^m + \eta \cdot (\pi_t - \pi_{ss}), \tag{3.6}$$

where $\eta > 1$ is the parameter that governs the response of nominal interest rate to inflationary pressures. In addition, i_∞^m is chosen to guarantee an inflation target π_{ss}.

Fiscal rule. To allow comparisons among policies, lump-sum transfers and credit policies, we setup a path for government debt as a policy target. During the crisis, we allow debt to expand and then shrink it back. We assume the following rule

$$\mathcal{E}_t = \underbrace{\mathcal{E}_\infty}_{\text{long-run target}} + \underbrace{(\mathcal{E}_d - \mathcal{E}_\infty) \cdot \exp\left[-\gamma^{LR} t\right]}_{\text{long-run deviation}} + \underbrace{(\mathcal{E}_{t^-} - \mathcal{E}_d) \cdot \exp\left[-\gamma^{SR} t\right]}_{\text{short-run deviation}}. \tag{3.7}$$

The term $\mathcal{E}_\infty$ is a long-run target. The term $\mathcal{E}_d$ is an attraction point of net asset position. The rate $\mathcal{E}_{t^-}$ is the government net asset position the instant before a shock. Finally, the

term $\exp\left(-\gamma^{SR}t\right)$ captures the speed of expansion of government debt and $\exp\left(-\gamma^{LR}t\right)$ the speed of reversal of the discretionary policy to the long-run target.We further discuss this rule in the Appendix A, where we argue that by assuming $\gamma^{LR} < \gamma^{SR}$, the net asset position (debt) decreases (increases) and then increases (decreases) during and after the crisis to accommodate the stabilizing policies. Since the path of net asset position is pinned down, we obtain the possible mix of transfers and credit policy as a residual as follows. Take derivatives with respect to time to obtain:

$$\dot{\mathcal{E}}_t = -\left[\gamma^{LR}\left(\mathcal{E}_d - \mathcal{E}_\infty\right)\cdot\exp\left(-\gamma^{LR}t\right) + \gamma^{SR}\left(\mathcal{E}_{t^-} - \mathcal{E}_d\right)\cdot\exp\left(-\gamma^{SR}t\right)\right]. \tag{3.8}$$

And thus, all policies that combine paths for transfers T_t and credit subsidy σ_t satisfying

$$T_t = r_t^m\mathcal{E}_t + \left(\sigma_t - \pi_t\right)\int_{\bar{s}}^{0} s\left[f\left(e,s,t\right) + f\left(u,s,t\right)\right]ds + \mathcal{W}_t\left(\tau^l\left(1-U_t\right) - bU_t\right) - \dot{\mathcal{E}}_t, \tag{3.9}$$

they also imply the same path for debt and, thus, they are comparable.

We consider two types of policy. First, a pure transfers policy, for which we set $\sigma_t - \pi_t = 0$, and thus solve for the path of lump-sum transfers that is consistent with the expansion of debt in (3.7). In this case,

$$T_t = r_t\mathcal{E}_t + \mathcal{W}_t\left(\tau^l\left(1-U_t\right) - bU_t\right) - \dot{\mathcal{E}_\sqcup}.$$

In addition, we consider an active credit policy for which we set the path credit subsidy according to:

$$\sigma_t = \underbrace{\sigma_\infty}_{\text{long-run target}} + \underbrace{\left(\sigma_d - \sigma_\infty\right)\cdot\exp\left[-\psi^{LR}t\right]}_{\text{long-run deviation}} + \underbrace{\left(\sigma_{t^-} - \sigma_d\right)\cdot\exp\left[-\psi^{SR}t\right]}_{\text{short-run deviation}}, \tag{3.10}$$

and transfers are backed as a residual from (3.9). This subsidy rule is akin to the evolution of debt in (3.7), in which the value σ_∞ plays to role of the long-run target. Again, the

term σ_d is an attraction point of the subsidy policy, and the rate $\sigma^m_{t^-}$ is the credit policy the instant before a shock. As regarding the law of motion to debt, the term $\exp\left(-\psi^{SR}t\right)$ captures a degree of responsiveness to the shock: the speed at which the discretionary policy kicks, whereas $\exp\left(-\psi^{LR}t\right)$ the speed of reversal of the discretionary policy, to the long-run target. Again, we further discuss this rule in the Appendix A. In this case, we assume that $\psi^{LR} > \psi^{SR}$ such that the credit subsidy σ_t first increases, and then decreases toward its long-run target.

One important aspect of macroeconomic stabilization policy is the speed of the implementation. The Covid-19 shock has evolved faster than ever. Arguably, a credit policy is faster to implement, given that monetary operations do not require the bureaucratic burden of a sending checks. We can control the speed of the policy responses via the parameters $\left\{\psi^{SR}, \gamma^{SR}\right\}$. Another critical aspect, is that we do not consider targeted transfers. Thus, the transfer program wastes budgetary resources distributing resources to wealthy individuals. Although in practice, transfers have been assigned to low-wage earners, the misallocation of transfers is still a problem. There are many low wage earners that have years of savings, and many high wage earners that are in deep debts. Without appropriately observing wealth, the effectiveness of the policy faces clear implementation constraints.

Finally, we could enrich the model to evaluate different schemes to finance those policies. For instance, by introducing short-run deviations of the Taylor rule that imply a lower path for the interest i^m_t, the debt evolution is mitigated, but this can potentially introduce trade-offs. Hence, the model can also provide insights on the welfare implications of the fiscal-monetary interactions needed to finance the debt implied by policies.

3.2.6 General equilibrium

Distribution of wealth and employment status. The mass of agents sums to one. At each instant t, there is a distribution $f(z, s, t)$ of real financial wealth across households given their employment status z. The law of motion of this distribution satisfies a Kolmogorov-

Forward Equation (KFE). The KFEs are given by:

$$\begin{aligned}\frac{\partial}{\partial t}f(e,s,t) &= -\frac{\partial}{\partial s}\left[\mu(e,s,t)f(e,s,t)\right]-\Gamma_t^{eu}f(e,s,t)+\Gamma_t^{ue}f(u,s,t) \text{ and}\\ \frac{\partial}{\partial t}f(u,s,t) &= -\frac{\partial}{\partial s}\left[\mu(u,s,t)f(u,s,t)\right]-\Gamma_t^{ue}f(u,s,t)+\Gamma_t^{eu}f(e,s,t). \end{aligned} \tag{3.11}$$

Note that a fraction $U_t = \int_{\bar{s}}^{\infty} uf(u,s,t)\,ds$ is unemployed, whereas a mass $1-U_t$ is active in the workforce. The mass of unemployed evolves according to the law of motion in (3.3).

Markets. Recall that $m_t^h(z,s)$, $a_t^h(z,s)$ and $l_t^h(z,s)$ are the demand for currency, deposits and loans, respectively, at instant t by a household with employment status z and savings s. Outside money is held as bank reserves or currency. The aggregate currency stock is

$$M0_t \equiv \int_{\bar{s}}^{\infty}\left[m_t^h(e,s)f(e,s,t)+m_t^h(u,s)f(u,s,t)\right]ds.$$

Naturally, households only hold currency at a zero-lower bound. Equilibrium in the outside money market is:

$$M0_t + M_t^b = M_t. \tag{3.12}$$

The credit market has two sides: a deposit and a loans market. In the deposit market, households hold deposits supplied by banks. In the loans market, households obtain loans supplied by banks and the CB. The distinction between the loans and deposits is that they clear with different interest rates. The deposit market clears when:

$$A_t^b = \int_0^{\infty}\left[a_t^h(e,s)f(e,s,t)+a_t^h(u,s)f(u,s,t)\right]ds, \tag{3.13}$$

whereas the loans market clears when:

$$L_t^b + L_t^f = \int_{\bar{s}}^{0}\left[l_t^h(e,s)f(e,s,t)+l_t^h(u,s)f(u,s,t)\right]ds, \tag{3.14}$$

whereL_t^b and L_t^f are loans purchases by banks and the CB, respectively. Note that household deposit demand is given by $a_t^h(z,s) \equiv P_t s - m_t^h(z,s)$ for positive values of s whereas demand for loans by $l_t^h(z,s) \equiv -P_t s$ for negative values of s.

Finally, the goods market clears when:

$$(1-U_t) \equiv Y_t = C_t \equiv \int_{\bar{s}}^{\infty} \sum_{z\in\{e,u\}} \left(c_t^r(z,s) + c_t^s(z,s)\right) f(z,s,t)\, ds, \tag{3.15}$$

where $c_t^r(z,s)$ and $c_t^s(z,s)$ represent the demands for remote and social goods, respectively, at instant t by a household with employment status z and savings s. The definition of the perfect foresight equilibrium is standard.

Equilibrium computation. The real equilibrium deposit rate solves a single clearing condition (Bigio and Sannikov, 2021):

$$-\int_{\bar{s}}^{0} s\left[f(e,s,t) + f(u,s,t)\right] ds = \int_{0}^{\infty} s\left[f(e,s,t) + f(u,s,t)\right] ds + \mathcal{E}_t \quad \text{for } t\in[0,\infty). \tag{3.16}$$

If we obtain the real deposit rate, we also obtain the real value of loans and deposits as well as the evolution of wealth. By Walras Law, if (3.16) holds, then the goods market clearing condition (3.15) also holds.

3.3 Policy responses and trade-offs

3.3.1 Two insights

We state here two insights to guide the discussion below. First, consider the case in which the borrowing limit is the natural one. Since this borrowing limit is never binding, and except for the initial unexpected shock there is not aggregate risk, the Ricardian equivalence holds meaning that debt and lump-sum transfers are equivalent to finance expenditures—see

Ljungqvist and Sargent (2012) for a textbook treatment. Hence, if the economy eventually returns to the same steady-state, any path of transfers is neutral as it will be consistent with a path in which transfers are set to zero all time. This insight showcases an extreme economy where one of the policies we evaluate has muted effects. Second, consider the other opposite case, an economy where debt limits are severe, as in Werning (2015) and Guerrieri et al. (2020). In particular, let the borrowing limit be zero. By construction, any credit subsidy is immaterial as there is not credit to be subsidized.

This discussion showcases that the effectiveness of policy critically depends on the extent of borrowing limits. In one extreme, when credit is ample, transfers are useless. In the other case, when credit is restricted, a credit policy is useless. These insights might be important for ongoing policy debates, as it means that in developed economies, with ample credit limits, transfer policies are more likely to to be neutral, whereas in emerging economies, with low credit limit, credit subsidy might not have a bite. These insights seem to be guiding policies to some extent: as we noted in the introduction, developing countries are relying more intensively on transfer programs, developed economies on credit programs.

3.3.2 Numerical illustrations

To evaluate the model, we present a simple calibration for illustrative purposes. As explained above, we assume that $\epsilon = 1.8$ and $\gamma = 0.5$, such that the shock in the marginal utility of social consumption is associated with less discounting, and thus, due to $IES > 1$ and rigid prices, it generates a recession. Also, we assume a low degree of substitutability, which is in line with the idea that some substitution occurs (e.g., going to the movies for online streaming), but given the nature of goods that requires social interactions, we doubt they can be largely substituted for goods that can be consumed at home.

We assume the economy is initially in the steady-state. Then, we solve the model for a time-varying path of β_t, which governs the evolution of behavioral responses due to the infection risk of Covid-19. Eventually, the economy converges back to steady-state. We

assume β_t follows an inverse-hump-shape path (see Figure 3.2, left-top panel), which becomes perfectly foreseen once the unforeseen Covid-19 shock hit the economy.

We consider a few exercises. First, we analyze the model dynamics in an environment with fully flexible prices. Second, we compare the model dynamics in an environment with rigid prices under three different scenarios: no policies at all (laissez-faire), a pure transfers policy in which the credit subsidy is set to zero, and an active credit subsidy in which lump-sum transfers are computed as a residual. Finally, we experiment with different degrees of borrowing limits. We consider the natural borrowing limit, in which transfers are immaterial due to arguments related to the Ricardian equivalence, so credit subsidy is the preferred policy. We also consider a zero borrowing limit, in which a credit subsidy by construction cannot do anything, so lump-sum transfers are the preferred one. And, finally, an intermediate borrowing limit in which trade-offs between the use of both policies emerge.

3.3.2.1 Benchmark: flexible prices

To set the stage, in this section, we report results under a flexible prices benchmark, which we obtain by setting the parameter κ to infinity. Figure 3.2 displays four plots. The top-left panel displays the evolution of β_t, the shock governing the marginal utility of consuming social goods. Under the assumption that $\epsilon > 1$ and $\gamma < 1$, this shock affects the economy as if there is less discounting of the future. Market clearing implies that consumption equals output, which is not affected by the shock since labor flows are exogenous. As the bottom panels reveal, output is constant and households simply reallocate consumption from social goods, affected by the negative marginal utility shock, toward remote goods. The price margin that adjusts is the real interest rate. Regarding the path of the interest rate, this path first increases at the time the shock arrives unexpectedly. Intuitively, given that households anticipate a fall in the the marginal utility of consumption of social goods as the shock escalates, they would like to anticipate consumption and, thus, interest rate must increase at the very beginning of the transition path. As the shock evolves in a perfect foreseen fashion,

interest rate follows an inverse hump shape, remaining temporarily below its steady-state level, given the incentives households face to postpone consumption to the future. Finally, given that output is constant, and consumption paths for social and remote goods depend mostly on γ and ϵ, the borrowing limit does not affect these paths. It affects only the path of interest rate which moves downwards as the borrowing limit becomes tighter and reduces the demand for loans.

This is a stark economy. Labor reallocation is perfect and instantaneous, meaning that a bar tender can become a UPS driver or a nurse the next day. This assumption is at odds with reality, but still, showcases that that some segments of the economy should absorb resources not employed in sectors that have to be avoided. The next sections address the points that with wage rigidity, remote sectors can be dragged down by the recession in social sectors.

3.3.2.2 Rigid prices, pure transfers and active credit policy

Under the assumption of price rigidity, $\kappa < \infty$, output drops in a persistent way, enhancing the role of potential policies to mitigate the impact of the shock. In the next sections, we consider scenarios without policy interventions (laissez-faire), with pure lump-sum transfers, and with an active credit policy where transfers adjust as a residual. To make policies comparable, we assume they generate the same path of the consolidated government's position in the latter two cases. We report results for three different degrees of borrowing limits. First, we consider the natural borrowing limit, in which transfers are innocuous. Second, we consider a zero borrowing limit in which credit subsidy is immaterial. Finally, an intermediate borrowing limit that illustrates the trade-off between using both policies.

Natural borrowing limit. Figure 3.3 considers the policy interventions, whereas Figure 3.4 displays output, consumption and interest rate paths. Compared to laissez-faire (full blue line), dashed red lines (dash-dotted green lines) represent the paths that take into account transfers policy (active credit policy) right after the shock, when the Government

instantaneously implements them. Policies are designed to generate the same path of debt, as illustrated in the first panel of Figure 3.3. The second and third panels plot paths of transfers and credit policy, respectively. In the later case, transfers adjust residually as illustrated by the seconde panel.

With nominal rigidity, and no policy interventions (blue full-lines in Figures 3.3 and 3.4), the output drops in a persistent way. As we explained earlier, this responds to the inability of the remote sector to absorb the slack of the social sector. There are more job separations which implies larger unemployment, and is captured by the slope of the output path. Micro uncertainty increases, meaning that precautionary savings should increase. This reinforces the negative effect on interest rates due to "less discounting" and $IES < 1$, or a more willingness to substitute intertemporally rather than intratemporally. Hence, as opposed to the flexible prices benchmark, consumption of remote goods does not absorb perfectly the fall in the consumption of social goods. There is scope for macroeconomic stabilization, a policy analysis we pursue next.

Due to arguments related to the Ricardian equivalence, except for approximation errors, the output, consumption and interest rate paths under laissez-faire and pure transfers overlap perfectly as portrayed in Figure 3.4. In contrast, if a credit policy is implemented by reducing the path of transfers, the fall in output is mitigated due to a boom in the consumption of remote goods. Due to the absence of any containment policy in our model, perhaps undesirably, the decline in the consumption of social goods is also mitigated. In addition, we do not see an initial increase in the interest rate any more, whose path follows an inverse hump shape, with the valley occurring before than in the laissez-faire economy. Intuitively, the incidence of a credit subsidy implies a lower rate that debtors must honor, but a higher rate that remunerates depositors. Hence, both paths of bank deposits and bank loans increase with respect to an economy where transfers are innocuous, as illustrated by Figure 3.5.

We do not analyze an economy with credit against future taxes. But the result is instructive: if borrowing limits are very large, then how much we vary the limit with transfers

will not have a large impact on the allocation.

Zero borrowing limit. Now we move to the opposite extreme. Consider Figures 3.6 and 3.7 that are the counterparts of Figures 3.4 and 3.5 for the case in which we assume no borrowing at all (we omit the paths of policies since they carry a similar message to the previous case in Figure 3.3). In the zero borrowing limit case, the credit policy is innocuous, and since transfers adjust residual to deliver the same expansion of debt as in the pure fiscal policy exercise, the paths under both policies (dashed red lines and dash-dotted green lines) overlap. This means that the marginal impact of the credit policy is muted here. Figure 3.6 reveals that, under the zero borrowing limit and laissez-faire, the same shock generates a recession nearly seven times larger than under the natural debt limit, which makes policy even more urgent. In other words, a stringent debt limit amplifies a lot the recession, at the same time that restrict the use of credit subsidy, perhaps the preferred policy tool to target those households who really need support. Once transfers kick in, the recession (as well as the fall in the consumption of social and remote goods) is mitigated substantially. As expected, the remote sector accommodates better the effect of transfers.

Under the zero borrowing limit and laissez-faire, interest rate initially increases abruptly as agents are willing to borrow to smooth the shock but the zero borrowing limit impedes so. Once transfers are implemented, as illustrated in Figure 3.7, part of it becomes deposits mitigating such increase in the interest rate.

Moderate borrowing limit. Finally, we consider the moderate borrowing limit. We calibrate $\bar{s}$ such that 10 percent of households are at the borrowing limit, $s_t = \bar{s}$, in the steady-state, which implies that 29 percent of them are indebted, $s_t \in [\bar{s}, 0)$. Figures 3.8 and 3.9 are the counterparts of Figures 3.4 and 3.5, and again we do not report the policy paths for conciseness. Recall that a pure fiscal policy (dashed red lines) means that transfers are set to match a given path of debt. As in the case with natural borrowing limit, under this

policy, the fall in output and consumption of both social and remote goods is mitigated with respect to the laissez-faire benchmark. Nonetheless, with an intermediate level of borrowing limit, if the government uses part of its fiscal resources to implement an active credit policy rather than solely pure transfers, it can further improves outcomes. Indeed, output and consumption paths move further upward (dashed-dot green lines), with a expansion in the remote good sector.

In addition, due to this mitigation in output and consumption once policies are implemented, interest rates do not increase as much as they do in the laissez-faire case to induce market clearing. The timing when rates fall below their steady-state level are anticipated once policies are implemented. With a moderate borrowing limit, transfers further increase deposits as part of the households save their transfers, but loans also fall as transfers to another part of the households weak precautionary needs at the same time that they can be used to anticipate consumption. Government debt absorbs such difference, affecting the path of interest rate in the aforementioned way. Under the credit policy, which stimulates both deposits and loans as illustrated by Figure 3.9, the interest rate increases less right after the shock, just to follow the same inverse hump-shape path towards the steady-state.

3.4 Final remarks

All the examples above are illustrative of the mechanism. The key message is that the best use of the mix between lump-sum transfers and active credit policy depends crucially on the extent of the borrowing limit. In the next versions of the paper, we aim to explore more this trade-off under a proper calibration and extensions that consider default risk and endogenize the borrowing limit. This is important to asses possible moral-hazard constraints. In addition, we aim to compare other types of policies, such as unemployment insurance and job retention schemes.

Next versions of the paper aim also to relax the assumption on perfect reallocation of

labor input across sectors, to study an ad-hoc restriction on the amount of social goods that can be consumed so we can simulate containment policies, and to allow temporary deviations from the Taylor rule so the consolidated government can tame the debt path with a lower interest rate.

3.A Appendix: Properties of policy rules

This section discusses the class of policy rules used in the draft. We specify rules of the form: :

$$x_t = x_\infty + (\bar{x}_0 - \bar{x}_\infty) \cdot \exp\left(-\mu^{LR} t\right) + (x_{0^-} - \bar{x}_0) \cdot \exp\left(-\mu^{SR} t\right).$$

In this rule, the value $\bar{x}_\infty$ is a long-run target. The term $\bar{x}_0$ is an attraction point of the policy rate in the short-run, after a shock. The rate x_{0^-} is the policy variable the instant before a shock. The term $\exp\left(-\mu^{SR} t\right)$ captures a degree of responsiveness to the shock to its short-run attraction point—the speed at which the discretionary policy kicks in, whereas $\exp\left(-\mu^{LR} t\right)$ the speed of reversal of the discretionary policy, to the long-run target. In what follows we assume $\mu^{LR} < \mu^{SR}$. This functional form has several natural properties:

1. First, observe that for any finite pair $\left\{\mu^{SR}, \mu^{LR}\right\}$ we have the following:

$$\lim_{t \to \infty} x_t = \bar{x}_\infty.$$

2. Consider that for any finite pair $\left\{\mu^{SR}, \mu^{LR}\right\}$ we have the following:

$$\lim_{t \to 0^+} x_t = x_{0^-}.$$

3. Consider that for any finite pair $\left\{\mu^{LR}\right\}$ we have the following:

$$\lim_{t \to 0^+} \lim_{\mu^{SR} \to \infty} x_t = \bar{x}_0,$$

   meaning that the adjustment is immediate.

4. Consider that for any finite pair $\left\{\mu^{SR}\right\}$ we have the following:

$$\lim_{t \to \infty} \lim_{\mu^{LR} \to \infty} x_t = \bar{x}_0,$$

meaning that the attraction point is the discretionary point.

5. Consider the limit, $\mu^{LR}/\mu^{SR} \to \infty$, then speed of responsiveness is immediate and the

$$\begin{aligned}
& \lim_{t\to\infty} \lim_{\mu^{SR}/\mu^{LR}\to\infty} x_t \\
= & \lim_{t\to\infty} \lim_{\mu^{SR}/\mu^{LR}\to\infty} \bar{x}_\infty + \exp\left(-\mu^{SR}t\right)\left[(\bar{x}_0 - \bar{x}_\infty)\cdot\exp\left(-\left(\mu^{LR}-\mu^{SR}\right)t\right) + (x_{0^-} - \bar{x}_0)\right] \\
= & \bar{x}_\infty + (\bar{x}_0 - \bar{x}_\infty) \lim_{t\to\infty} \lim_{\mu^{SR}/\mu^{LR}\to\infty} \exp\left(-\mu^{SR}t\right)\left[\exp\left(-\left(\mu^{LR}-\mu^{SR}\right)t\right)\right] \\
= & \bar{x}_\infty + (\bar{x}_0 - \bar{x}_\infty) \lim_{t\to\infty} \lim_{\mu^{SR}/\mu^{LR}\to\infty} \exp\left(-\mu^{SR}t\right)\left[\exp\left(\mu^{SR}\left(1-\mu^{LR}/\mu^{SR}\right)t\right)\right] \\
= & \bar{x}_0,
\end{aligned}$$

where the last line follows by L'Hospital rule.

6. Monotonicity of x_t. Let's assume $x_\infty > \bar{x}_0$ and $x_{0^-} > \bar{x}_0$, which is the scenario in our simulations. If $\mu^{LR} < \mu^{SR}$, then

$$\frac{\partial x_t}{\partial t} \gtreqless 0 \text{ iff } t \gtreqless \frac{1}{\mu^{SR}-\mu^{LR}} \ln\left(\frac{\mu^{SR}}{\mu^{LR}}\cdot\frac{x_{0^-}-\bar{x}_0}{x_\infty-\bar{x}_0}\right),$$

which means that the path of $\bar{x}_t$ first decreases over time from x_{0^-}, then increases back to x_∞, which is our parametrization for the evolution of real net position, $\mathcal{E}_t$, after the shock to represent the fiscal space available for stabilizing policies. If instead, $\mu^{LR} > \mu^{SR}$, then

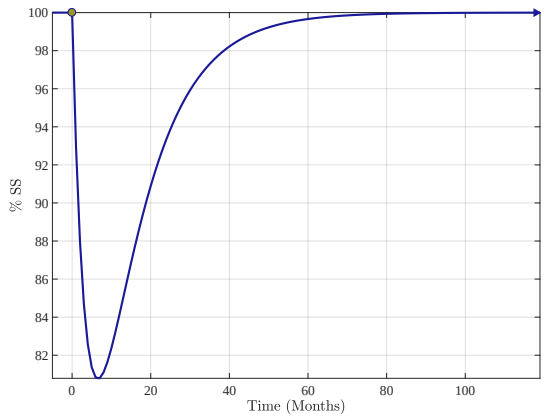
$$\frac{\partial \bar{x}_t}{\partial t} \gtreqless 0 \text{ iff } t \lesseqgtr \frac{1}{\mu^{LR}-\mu^{SR}} \ln\left(\frac{\mu^{LR}}{\mu^{SR}}\cdot\frac{x_\infty-\bar{x}_0}{x_{0^-}-\bar{x}_0}\right),$$

which means that the path of $\bar{x}_t$ first increases over time from x_{0^-}, then decreases back to $\bar{x}_\infty$, which is our choice whenever we study an active credit subsidy σ_t.

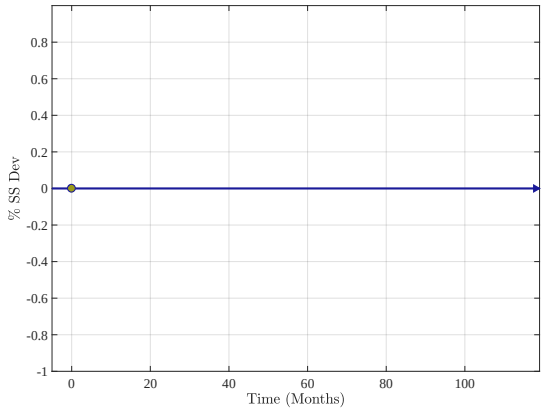
3.B Appendix: Figures

Figure 3.2: Transition paths under flexible prices

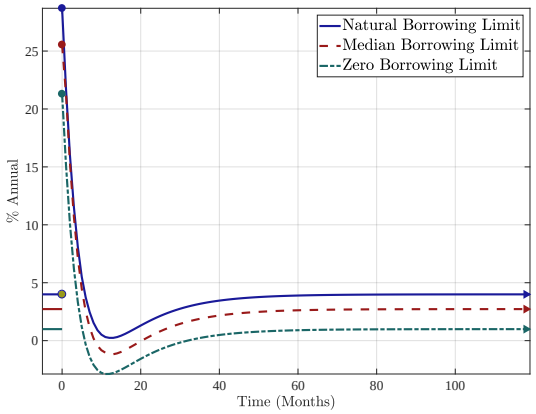
(a) Preference Shock $\beta(t)$



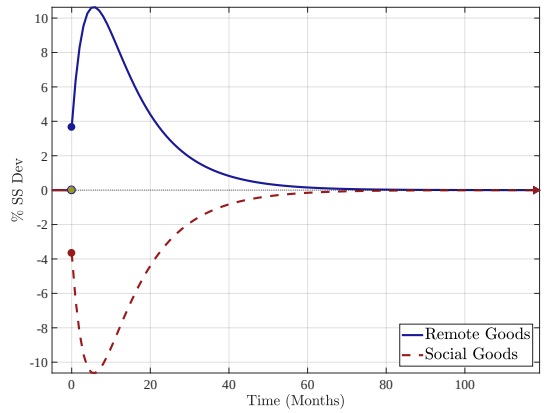
(b) Output Y_t



(c) Real Interest Rates



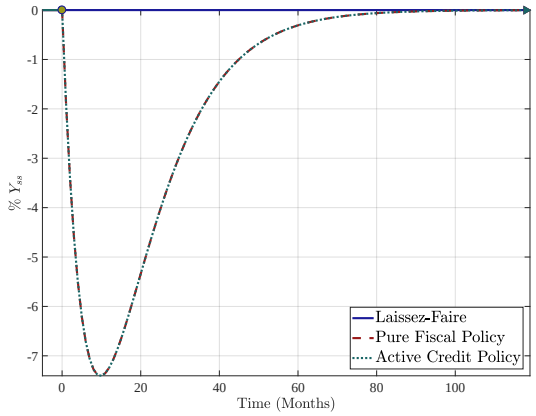
(d) Consumption C_t^r and C_t^s



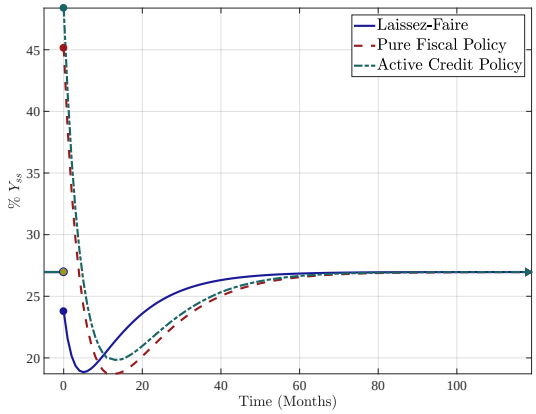
Notes: The figure reports the paths of preference shock, total output, real interest rate and consumption of social goods and remote goods under flexible prices after an unforeseen Covid-19 shock. In panel (a), the preference parameter of social goods consumption β_t is expressed in the percentage of steady-state value. In panels (b) and (d), the total output and consumption of social goods and remote goods are expressed in percentage deviations from the steady-state values. In panels (c), the real interest rates are expressed in annual percentages under three levels of borrowing limit: natural borrowing limit, zero borrowing limit, and a moderate borrowing limit where $\bar{s} = -0.1b$.

Figure 3.3: Nominal rigidity and policy variables (natural borrowing limit).

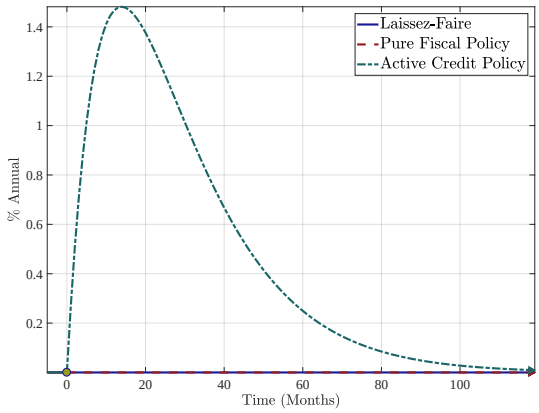
(a) Real Net Asset Position $\mathcal{E}_t$



(b) Real Fiscal Transfer



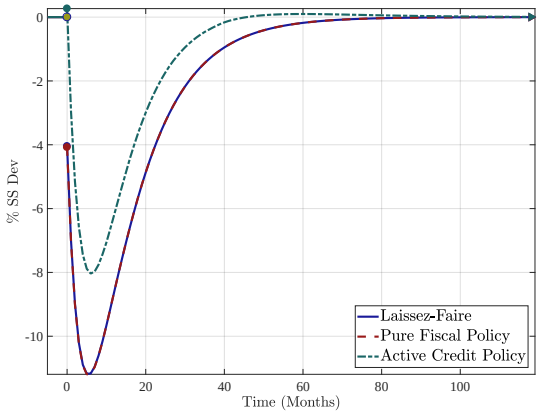
(c) Real Credit Subsidy $\sigma_t - \pi_t$



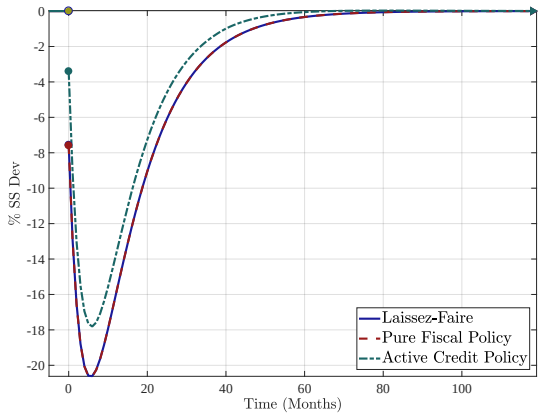
Notes: The figure reports the paths of real net asset position, real fiscal transfer and real credit subsidy in the following scenarios of policy interventions after an unforeseen Covid-19 shock: laissez-faire, pure lump-sum trasnfer and active credit policy. In panels (a) and (b), the net asset position and fiscal transfers are expressed in percentage terms of the steady-state output. In panel (c), the real credit subsidity rate is expressed in annual percentage. The pure fiscal transfer policy and active credit policy follow the same path of net asset position, and the credit subsidy rates in laissez-faire and pure fiscal transfer policy are equal to zero. Given the paths of net asset position and credit subsidy rate, the paths of fiscal transfers are computed as residuals from equation (3.9). In all figures the households' borrowing limit is set at the natural borrowing limit level, such that the mass of population constrained at the limit is zero along the whole path of transition.

Figure 3.4: Nominal rigidity and real variables (natural borrowing limit).

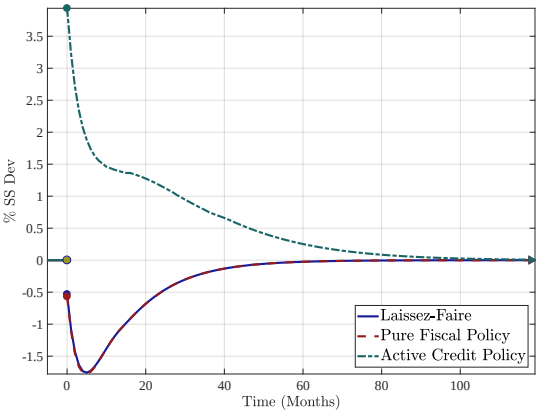
(a) Output Y_t



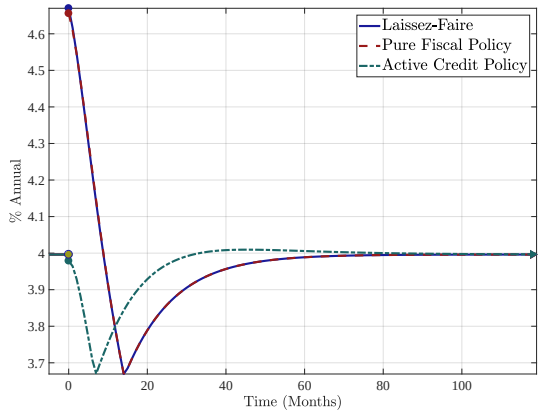
(b) Consumption of Social Goods C_t^s



(c) Consumption of Remote Goods C_t^r



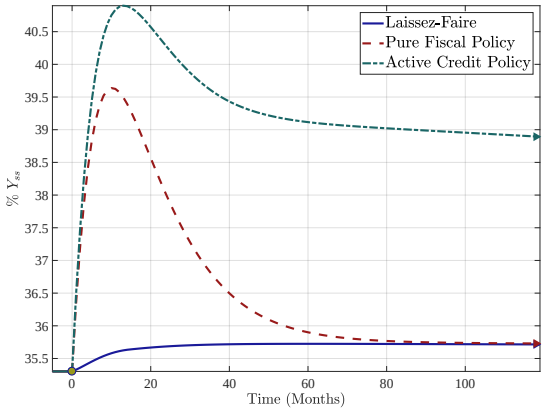
(d) Real Interest Rates



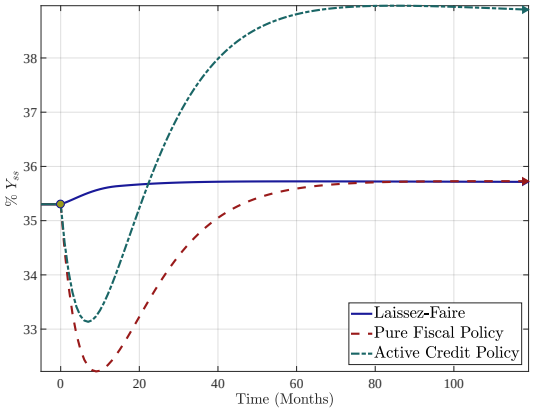
Notes: The figure reports the paths of total output, consumption of social goods and remote goods, and real interest rate in the following scenarios of policy interventions after an unforeseen Covid-19 shock: laissez-faire, pure lump-sum trasnfer and active credit policy. In panels (a), (b) and (c), the total output and consumption of social and remote goods are expressed in percentage deviations from the steady-state values. In panel (d), the real interest rates are expressed in annual percentages. In all figures the households' borrowing limit is set at the natural borrowing limit level, such that the mass of population constrained at the limit is zero along the whole path of transition. The paths of policy interventions follow Figure 3.3.

Figure 3.5: Nominal rigidity and banking variables (natural borrowing limit).

(a) Bank Deposits A_t



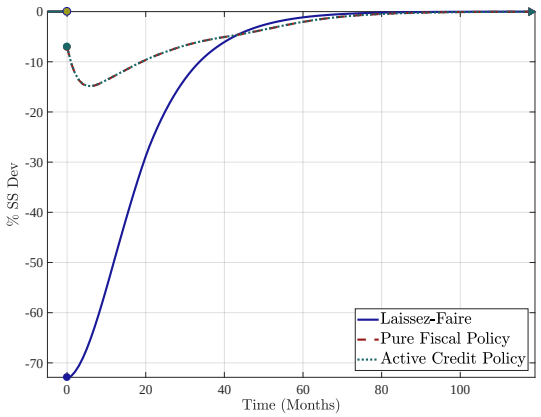
(b) Bank Loans L_t



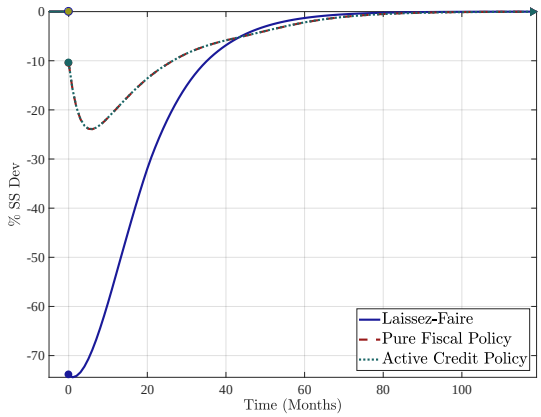
Notes: The figure reports the paths of bank deposits and loans in the following scenarios of policy interventions after an unforeseen Covid-19 shock: laissez-faire, pure lump-sum trasnfer and active credit policy. The bank deposits and loans are expressed in percentages of the steady-state output. In all figures the households' borrowing limit is set at the natural borrowing limit level, such that the mass of population constrained at the limit is zero along the whole path of transition. The paths of policy interventions follow Figure 3.3.

Figure 3.6: Nominal rigidity and real variables (zero borrowing limit).

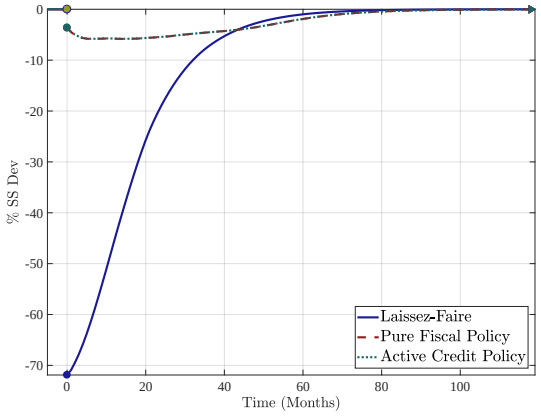
(a) Output Y_t



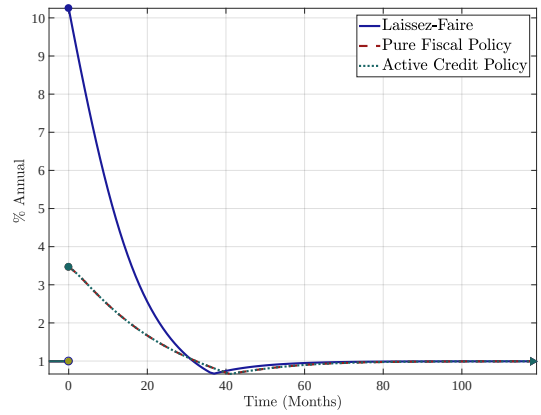
(b) Consumption of Social Goods C_t^s



(c) Consumption of Remote Goods C_t^r



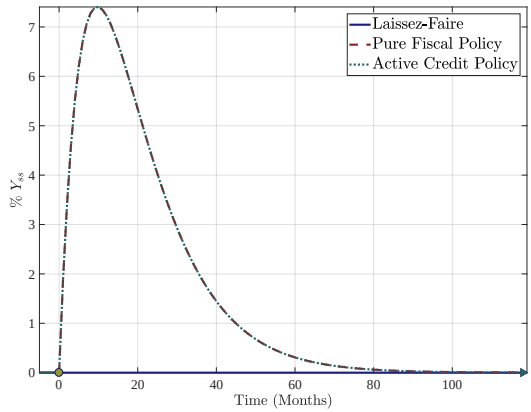
(d) Real Deposit Rate



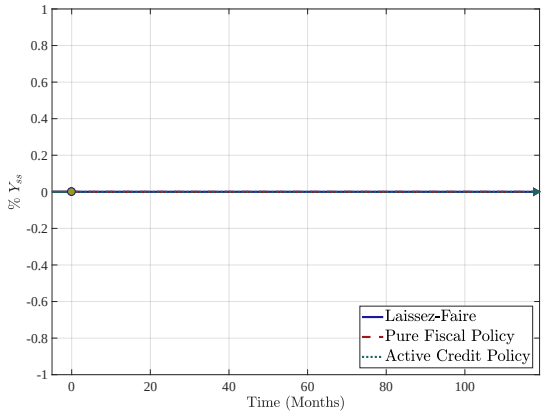
Notes: The figure reports the paths of total output, consumption of social goods and remote goods, and real interest rate in the following scenarios of policy interventions after an unforeseen Covid-19 shock: laissez-faire, pure lump-sum trasnfer and active credit policy. In panels (a), (b) and (c), the total output and consumption of social and remote goods are expressed in percentage deviations from the steady-state values. In panel (d), the real interest rates are expressed in annual percentages. In all figures the households' borrowing limit is set at the zero borrowing limit level, i.e., $\bar{s} = 0$. The paths of net asset position and credit subsidy rate follow Figure 3.3 and the paths of lump-sum transfers are computed as residuals from equation (3.9).

Figure 3.7: Nominal rigidity and banking variables (zero borrowing limit).

(a) Bank Deposits A_t



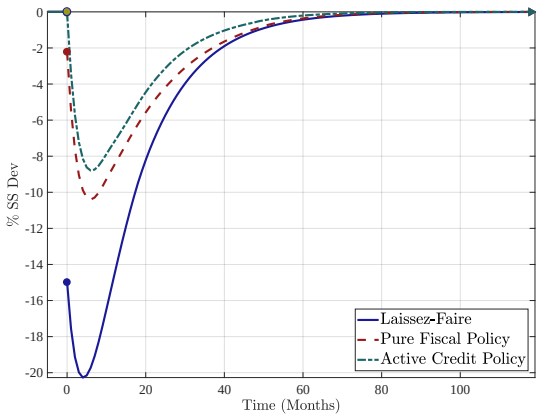
(b) Bank Loans L_t



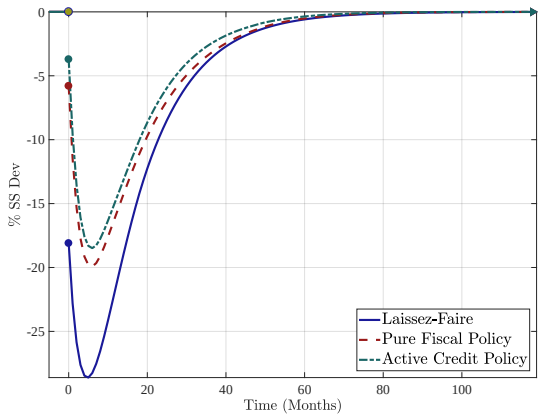
Notes: The figure reports the paths of bank deposits and loans in the following scenarios of policy interventions after an unforeseen Covid-19 shock: laissez-faire, pure lump-sum trasnfer and active credit policy. The bank deposits and loans are expressed in percentages of the steady-state output. In all figures the households' borrowing limit is set at the zero borrowing limit level, i.e., $\bar{s} = 0$. The paths of net asset position and credit subsidy rate follow Figure 3.3, and the paths of lump-sum transfers are computed as residuals from equation (3.9).

Figure 3.8: Nominal rigidity and real variables (moderate borrowing limit).

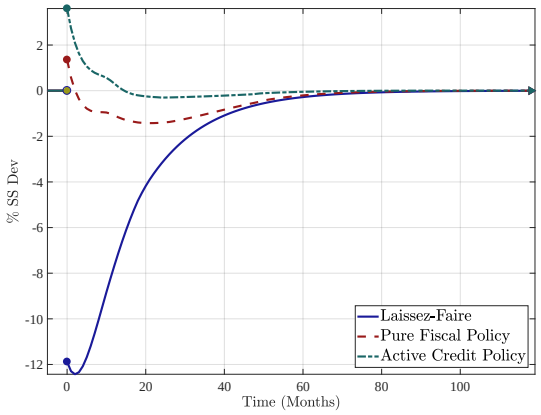
(a) Output Y_t



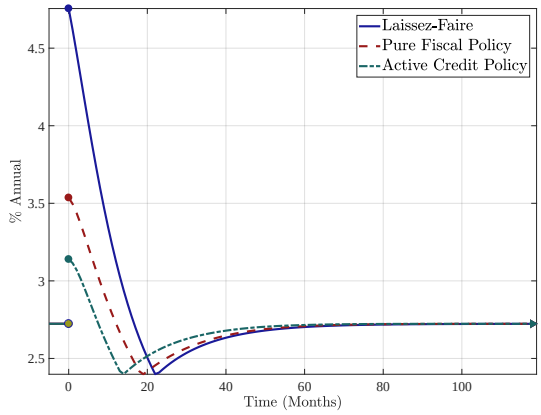
(b) Consumption of Social Goods C_t^s



(c) Consumption of Remote Goods C_t^r



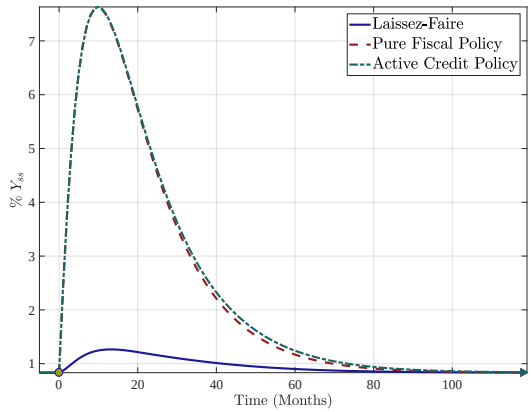
(d) Real Deposit Rate



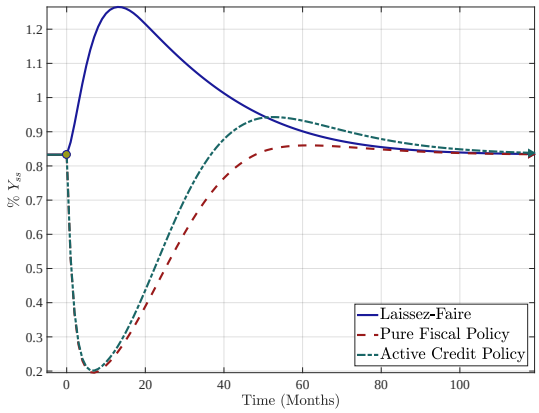
Notes: The figure reports the paths of total output, consumption of social goods and remote goods, and real interest rate in the following scenarios of policy interventions after an unforeseen Covid-19 shock: laissez-faire, pure lump-sum trasnfer and active credit policy. In panels (a), (b) and (c), the total output and consumption of social and remote goods are expressed in percentage deviations from the steady-state values. In panel (d), the real interest rates are expressed in annual percentages. In all figures the households' borrowing limit is set at a moderate level, i.e., $\bar{s} = -0.1b$. The paths of net asset position and credit subsidy rate follow Figure 3.3, and the paths of lump-sum transfers are computed as residuals from equation (3.9).

Figure 3.9: Nominal rigidity and banking variables (moderate borrowing limit).

(a) Bank Deposits A_t



(b) Bank Loans L_t



Notes: The figure reports the paths of bank deposits and loans in the following scenarios of policy interventions after an unforeseen Covid-19 shock: laissez-faire, pure lump-sum trasnfer and active credit policy. The bank deposits and loans are expressed in percentages of the steady-state output. In all figures the households' borrowing limit is set at a moderate level, i.e., $\bar{s} = -0.1b$. The paths of net asset position and credit subsidy rate follow Figure 3.3, and the paths of lump-sum transfers are computed as residuals from equation (3.9).

Bibliography

Acosta, M. and Saia, J. (2020). Estimating the effects of monetary policy via high frequency factors. Columbia University working paper. 111

Adda, J. and Cooper, R. (2003). *Dynamic economics: quantitative methods and applications*. MIT press. 230

Afonso, G., Armenter, R., and Lester, B. (2019). A model of the federal funds market: yesterday, today, and tomorrow. *Review of Economic Dynamics*, 33:177–204. 31, 100

Afonso, G., Kovner, A., and Schoar, A. (2011). Stressed, not frozen: The federal funds market in the financial crisis. *Journal of Finance*, 66(4):1109–1139. 157, 226

Afonso, G. and Lagos, R. (2015a). The over-the-counter theory of the fed funds market: A primer. *Journal of Money, Credit and Banking*, 47(S2):127–154. 100

Afonso, G. and Lagos, R. (2015b). Trade dynamics in the market for federal funds. *Econometrica*, 83(1):263–313. iii, 97, 98, 100, 108, 113, 116, 119, 124, 127, 142, 152

Aguirregabiria, V., Clark, R., and Wang, H. (2019). The geographic flow of bank funding and access to credit: Branch networks, local synergies, and competition. CEPR Discussion Paper No. DP13741. 33

Alvarez, F., Argente, D., and Lippi, F. (forthcoming). A simple planning problem for covid-19 lockdown, testing, and tracing. *American Economic Review: Insights*. 236

Armenter, R. and Lester, B. (2017). Excess reserves and monetary policy implementation. *Review of Economic Dynamics*, 23:212–235. 31

Atkeson, A. (2020). What will be the economic impact of covid-19 in the us? rough estimates of disease scenarios. Working Paper 26867, National Bureau of Economic Research. 236

Balloch, C. and Koby, Y. (2019). Low rates and bank loan supply: Theory and evidence from japan. Technical report, Working paper, London School of Economics. 1, 7, 9

Barro, R. J., Ursúa, J. F., and Weng, J. (2020). The coronavirus and the great influenza pandemic: Lessons from the "spanish flu" for the coronavirus's potential effects on mortality and economic activity. Working Paper 26866, National Bureau of Economic Research. 236

Bech, M. and Keister, T. (2017). Liquidity regulation and the implementation of monetary policy. *Journal of Monetary Economics*, 92:64–77. 100

Bech, M. and Monnet, C. (2016). A search-based model of the interbank money market and monetary policy implementation. *Journal of Economic Theory*, 164:32–67. 100

Bech, M. L. and Atalay, E. (2010). The topology of the federal funds market. *Physica A: Statistical Mechanics and its Applications*, 389(22):5223–5246. 100

Bech, M. L. and Malkhozov, A. (2016). How have central banks implemented negative policy rates? *BIS Quarterly Review March*. 1

Begenau, J., Bigio, S., Majerovitz, J., and Vieyra, M. (2020). A q-theory of banks. Working Paper 27935, National Bureau of Economic Research. 23

Begenau, J. and Stafford, E. (2019). Do banks have an edge? working paper, available at SSRN 3095550. 31

Bemanke, B. and Gertler, M. (1989). Agency costs, net worth, and business fluctuations. *American Economic Review*, 79(1):14–31. 7

Benhabib, J. and Farmer, R. E. (1994). Indeterminacy and increasing returns. *Journal of Economic Theory*, 63(1):19–41. 120

Berentsen, A. and Monnet, C. (2008). Monetary policy in a channel system. *Journal of Monetary Economics*, 55(6):1067–1080. 100

Berger, A. N. and Hannan, T. H. (1989). The price-concentration relationship in banking. *Review of Economics and Statistics*, pages 291–299. 8

Berger, D. W., Herkenhoff, K. F., and Mongey, S. (2020). An seir infectious disease model with testing and conditional quarantine. Working Paper 26901, National Bureau of Economic Research. 236

Bernanke, B. S. (1983). Nonmonetary effects of the financial crisis in the propagation of the great depression. *American Economic Review*, 73(3):257–276. 6

Bernanke, B. S. and Blinder, A. S. (1988). Credit, money, and aggregate demand. *American Economic Review*, 78(2):435–439. 6

Bernanke, B. S. and Blinder, A. S. (1992). The federal funds rate and the channels of monetary transmission. *American Economic Review*, 82(4):901–921. 6

Berriel, T. and Zilberman, E. (2011). Targeting the poor: A macroeconomic analysis of cash transfer programs. PUC-Rio working paper. 235

Bethune, Z. A. and Korinek, A. (2020). Covid-19 infection externalities: Trading off lives vs. livelihoods. Working Paper 27009, National Bureau of Economic Research. 236

Bianchi, J. and Bigio, S. (forthcoming). Banks, liquidity management and monetary policy. *Econometrica*. 100

Bigio, S. and Sannikov, Y. (2021). A model of credit, money, interest, and prices. Working Paper 28540, National Bureau of Economic Research. 7, 100, 233, 251

Bindseil, U. (2018). *Financial stability implications of a prolonged period of low interest rates*. BIS. 1

Bolton, P. and Freixas, X. (2000). Equity, bonds, and bank debt: Capital structure and financial market equilibrium under asymmetric information. *Journal of Political Economy*, 108(2):324–351. 7

Bracewell, R. N. (2000). *The Fourier transform and its applications*. McGraw-Hill New York. 186

Brunnermeier, M., Landau, J.-P., Pagano, M., and Reis, R. (2020). Throwing a covid-19 liquidity life-line. *Economics for Inclusive Prosperity (Econfip)*. 231

Brunnermeier, M. K. and Koby, Y. (2018). The reversal interest rate. Working Paper 25406, National Bureau of Economic Research. 1, 7, 23

Brunnermeier, M. K. and Sannikov, Y. (2014). A macroeconomic model with a financial sector. *American Economic Review*, 104(2):379–421. 7

Brunnermeier, M. K. and Sannikov, Y. (2016). The i theory of money. Working Paper 22533, National Bureau of Economic Research. 7

Buera, F. J., Fattal-Jaef, R. N., Hopenhayn, H., Neumeyer, P. A., and Shin, Y. (2021). The economic ripple effects of covid-19. Working Paper 28704, National Bureau of Economic Research. 236

Caballero, R. J. and Simsek, A. (forthcoming). A model of endogenous risk intolerance and lsaps: Asset prices and aggregate demand in a "covid-19" shock. *Review of Financial Studies*. 236

Chang, B. and Zhang, S. (2018). Endogenous market making and network formation. working paper, available at SSRN 2600242. 101

Chen, B. S., Hanson, S. G., and Stein, J. C. (2017). The decline of big-bank lending to small business: Dynamic impacts on local credit and labor markets. Working Paper 23843, National Bureau of Economic Research. 33

Chetty, V. K. (1969). On measuring the nearness of near-moneys. *American Economic Review*, 59(3):270–281. 22

Chiu, J., Eisenschmidt, J., and Monnet, C. (2020). Relationships in the interbank market. *Review of Economic Dynamics*, 35:170–191. 100

Claessens, S., Coleman, N., and Donnelly, M. (2018). "low-for-long" interest rates and banks' interest margins and profitability: Cross-country evidence. *Journal of Financial Intermediation*, 35:1–16. 1

Corbae, D. and D'Erasmo, P. (forthcoming). Capital buffers in a quantitative model of banking industry dynamics. *Econometrica*. 8

Corbae, D. and Levine, R. (2019). Competition, stability, and efficiency in the banking industry. Manuscript, University of Wisconsin. 8

Correia, S., Luck, S., and Verner, E. (2020). Pandemics depress the economy, public health interventions do not: Evidence from the 1918 flu. working paper. 236

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *Quarterly Journal of Economics*, 135(2):561–644. 8

Decker, R., Haltiwanger, J., Jarmin, R., and Miranda, J. (2014). The role of entrepreneurship in us job creation and economic dynamism. *Journal of Economic Perspectives*, 28(3):3–24. 47

Di Tella, S. and Kurlat, P. (forthcoming). Why are banks exposed to monetary policy? *American Economic Journal: Macroeconomics*. 22

Diebold, F. X. and Sharpe, S. A. (1990). Post-deregulation bank-deposit-rate pricing: The multivariate dynamics. *Journal of Business & Economic Statistics*, 8(3):281–291. 8

Diez, M. F., Leigh, M. D., and Tambunlertchai, S. (2018). *Global market power and its macroeconomic implications*. International Monetary Fund. 8

Drechsler, I., Savov, A., and Schnabl, P. (2017). The deposits channel of monetary policy. *Quarterly Journal of Economics*, 132(4):1819–1876. 4, 7, 8, 9, 12, 22, 29, 33, 36, 40, 47, 50

Drechsler, I., Savov, A., and Schnabl, P. (2019). How monetary policy shaped the housing boom. Working Paper 25649, National Bureau of Economic Research. 4

Drechsler, I., Savov, A., and Schnabl, P. (forthcoming). Banking on deposits: Maturity transformation without interest rate risk. *Journal of Finance*. 9, 31

Driscoll, J. C. and Judson, R. (2013). Sticky deposit rates. working paper, available at SSRN 2241531. 8

Duffie, D., Gârleanu, N., and Pedersen, L. H. (2005). Over-the-counter markets. *Econometrica*, 73(6):1815–1847. 100

Duffie, D. and Krishnamurthy, A. (2016). Passthrough efficiency in the fed's new monetary policy setting. In *Designing Resilient Monetary Policy Frameworks for the Future. Federal Reserve Bank of Kansas City, Jackson Hole Symposium*, pages 1815–1847. 8, 31, 100

Eggertsson, G. B., Juelsrud, R. E., Summers, L. H., and Wold, E. G. (2019). Negative nominal interest rates and the bank lending channel. Working Paper 25416, National Bureau of Economic Research. 7

Eggertsson, G. B. and Woodford, M. (2006). Optimal Monetary and Fiscal Policy in a Liquidity Trap. In *NBER International Seminar on Macroeconomics 2004*, NBER Chapters, pages 75–144. National Bureau of Economic Research, Inc. 7

Eichenbaum, M. S., Rebelo, S., and Trabandt, M. (2020). The macroeconomics of epidemics. Working Paper 26882, National Bureau of Economic Research. 236

English, W. B., Van den Heuvel, S. J., and Zakrajšek, E. (2018). Interest rate risk and bank equity valuations. *Journal of Monetary Economics*, 98:80–97. 31, 34, 42

Ennis, H. M. (2018). A simple general equilibrium model of large excess reserves. *Journal of Monetary Economics*, 98:50–65. 100

Farboodi, M., Jarosch, G., and Shimer, R. (2017). The emergence of market structure. Working Paper 23234, National Bureau of Economic Research. 100, 146

Faria-e Castro, M. (2021). Fiscal policy during a pandemic. *Journal of Economic Dynamics and Control*, 125:104088. 236

Feenstra, R. C. (1986). Functional equivalence between liquidity costs and the utility of money. *Journal of Monetary Economics*, 17(2):271–291. 22

Fuster, A., Hizmo, A., Lambie-Hanson, L., Vickery, J., and Willen, P. (2021). How resilient is mortgage credit supply? evidence from the covid-19 pandemic. CEPR Discussion Paper No. DP16110. 23

Gertler, M. and Kiyotaki, N. (2010). Financial intermediation and credit policy in business cycle analysis. In *Handbook of monetary economics*, volume 3, pages 547–599. Elsevier. 7, 23

Gilje, E. P., Loutskina, E., and Strahan, P. E. (2016). Exporting liquidity: Branch banking and financial integration. *Journal of Finance*, 71(3):1159–1184. 36

Gofman, M. (2017). Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *Journal of Financial Economics*, 124(1):113–146. 100

Gomez, M., Landier, A., Sraer, D., and Thesmar, D. (2021). Banks' exposure to interest rate risk and the transmission of monetary policy. *Journal of Monetary Economics*, 117:543–570. 47

Gourinchas, P.-O. (2020). Flattening the pandemic and recession curves. *Mitigating the COVID Economic Crisis: Act Fast and Do Whatever*, 31:57–62. 231

Greenstone, M., Mas, A., and Nguyen, H.-L. (2020). Do credit market shocks affect the real economy? quasi-experimental evidence from the great recession and" normal" economic times. *American Economic Journal: Economic Policy*, 12(1):200–225. 33

Greenwald, D. L., Krainer, J., and Paul, P. (2020). The credit line channel. Federal Reserve Bank of San Francisco working paper. 47

Guerrieri, V., Lorenzoni, G., Straub, L., and Werning, I. (2020). Macroeconomic implications of covid-19: Can negative supply shocks cause demand shortages? Working Paper 26918, National Bureau of Economic Research. 233, 234, 236, 239, 252

Hamilton, J. D. (1996). The daily market for federal funds. *Journal of Political Economy*, 104(1):26–56. iii, 100, 152

Hannan, T. H. and Berger, A. N. (1991). The rigidity of prices: Evidence from the banking industry. *American Economic Review*, 81(4):938–945. 8

He, Z. and Krishnamurthy, A. (2013). Intermediary asset pricing. *American Economic Review*, 103(2):732–70. 7

Holmstrom, B. and Tirole, J. (1997). Financial intermediation, loanable funds, and the real sector. *Quarterly Journal of economics*, 112(3):663–691. 23

Hugonnier, J., Lester, B., and Weill, P.-O. (2020). Frictional intermediation in over-the-counter markets. *Review of Economic Studies*, 87(3):1432–1469. 100

Ihrig, J. E., Vojtech, C. M., and Weinbach, G. C. (2019). How have banks been managing the composition of high-quality liquid assets? *Review*, 101(3):177–201. 157

Jackson, H. (2015). The international experience with negative policy rates. Technical report, Bank of Canada. 1

Jarociński, M. and Karadi, P. (2020). Deconstructing monetary policy surprises—the role of information shocks. *American Economic Journal: Macroeconomics*, 12(2):1–43. 30, 45, 47, 48, 71, 73, 74, 75

Jiménez, G., Ongena, S., Peydró, J.-L., and Saurina, J. (2014). Hazardous times for monetary policy: What do twenty-three million bank loans say about the effects of monetary policy on credit risk-taking? *Econometrica*, 82(2):463–505. 6

Jones, C. J., Philippon, T., and Venkateswaran, V. (forthcoming). Optimal mitigation policies in a pandemic: Social distancing and working from home. *Review of Financial Studies*. 236

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American economic review*, 95(1):161–182. 45, 66, 68, 69, 70, 71, 73, 74, 75

Jordà, O., Singh, S. R., and Taylor, A. M. (2020). Longer-run economic consequences of pandemics. Working Paper 26934, National Bureau of Economic Research. 236

Kaplan, G., Moll, B., and Violante, G. L. (2020). The great lockdown and the big stimulus: Tracing the pandemic possibility frontier for the u.s. Working Paper 27794, National Bureau of Economic Research. 236

Kashyap, A. K. and Stein, J. C. (1994). Monetary policy and bank lending. In *Monetary policy*, pages 221–261. The University of Chicago Press. 6

Kashyap, A. K. and Stein, J. C. (1995). The impact of monetary policy on bank balance sheets. In *Carnegie-Rochester conference series on public policy*, volume 42, pages 151–195. Elsevier. 6

Kashyap, A. K. and Stein, J. C. (2000). What do a million observations on banks say about the transmission of monetary policy? *American Economic Review*, 90(3):407–428. 6

Kashyap, A. K. and Stein, J. C. (2012). The optimal conduct of monetary policy with interest on reserves. *American Economic Journal: Macroeconomics*, 4(1):266–82. 100

Kashyap, A. K., Stein, J. C., and Wilcox, D. W. (1993). Monetary policy and credit conditions: Evidence from the composition of external finance. *American Economic Review*, 83(1):78–98. 6

Keating, T. and Macchiavelli, M. (2017). Interest on reserves and arbitrage in post-crisis money markets. FEDS Working Paper No. 2017-124. 105

Kiyotaki, N. and Moore, J. (1997). Credit cycles. *Journal of political economy*, 105(2):211–248. 7

Krueger, D., Uhlig, H., and Xie, T. (2020). Macroeconomic dynamics and reallocation in an epidemic: Evaluating the "swedish solution". Working Paper 27047, National Bureau of Economic Research. 236

Krugman, P. R., Dominquez, K. M., and Rogoff, K. (1998). It's baaack: Japan's slump and the return of the liquidity trap. *Brookings Papers on Economic Activity*, 1998(2):137–205. 7

Lagos, R. and Rocheteau, G. (2007). Search in asset markets: Market structure, liquidity, and welfare. *American Economic Review*, 97(2):198–202. 220

Lagos, R. and Rocheteau, G. (2009). Liquidity in asset markets with search frictions. *Econometrica*, 77(2):403–426. 100

Lagos, R. and Zhang, S. (2019). A monetary model of bilateral over-the-counter markets. *Review of Economic Dynamics*, 33:205–227. 100

Li, W., Ma, Y., and Zhao, Y. (2019). The passthrough of treasury supply to bank deposit funding. *Columbia Business School Research Paper*. 4, 12

Liu, S. (2020). Dealers' search intensity in us corporate bond markets. working paper, available at SSRN 3644132. 100

Ljungqvist, L. and Sargent, T. J. (2012). *Recursive Macroeconomic Theory*. MIT Press, Boston, 3rd edition edition. 252

Mankart, J., Michaelides, A., and Pagratis, S. (2020). Bank capital buffers in a dynamic model. *Financial Management*, 49(2):473–502. 32

Milgrom, P. and Shannon, C. (1994). Monotone comparative statics. *Econometrica*, 62(1):157–180. 168

Nagel, S. (2016). The liquidity premium of near-money assets. *Quarterly Journal of Economics*, 131(4):1927–1971. 22

Nakamura, E. and Steinsson, J. (2018). High-frequency identification of monetary non-neutrality: the information effect. *Quarterly Journal of Economics*, 133(3):1283–1330. 9, 11, 30, 45, 46, 47, 48, 66, 68, 69, 70, 111

Neumark, D. and Sharpe, S. A. (1992). Market structure and the nature of price rigidity: evidence from the market for consumer deposits. *Quarterly Journal of Economics*, 107(2):657–680. 8

Piazzesi, M. and Schneider, M. (2018). Payments, credit and asset prices. BIS Working Paper No. 734. 23

Poole, W. (1968). Commercial bank reserve management in a stochastic model: implications for monetary policy. *Journal of finance*, 23(5):769–791. 100, 224

Poterba, J. and Rottemberg, J. (1987). Money in the utility function: An empirical implementation, new approaches to monetary economics: Proceedings of the second international symposium in economic theory and econometrics. 22

Romer, C. D. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, 94(4):1055–1084. 30

Sá, A. I. and Jorge, J. (2019). Does the deposits channel work under a low interest rate environment? *Economics Letters*, 185:108736. 7

Scharfstein, D. and Sunderam, A. (2016). Market power in mortgage lending and the transmission of monetary policy. working paper. Harvard University. 7, 23, 33

Sigman, K. (2007). Poisson processes, and compound (batch) poisson processes. Lecture notes, Columbia University, http://www.columbia.edu/ ks20/4703-Sigman/4703-07-Notes-PP-NSPP.pdf. 228

Trejos, A. and Wright, R. (2016). Search-based models of money and finance: An integrated approach. *Journal of Economic Theory*, 164:10–31. 100

Ulate, M. (2021). Going negative at the zero lower bound: The effects of negative nominal interest rates. *American Economic Review*, 111(1):1–40. 1, 7, 23

Üslü, S. (2019). Pricing and liquidity in decentralized asset markets. *Econometrica*, 87(6):2079–2140. 100, 124, 128, 173

Van den Heuvel, S. J. (2002). The bank capital channel of monetary policy. The Wharton School, University of Pennsylvania, mimeo, 2013–14. 7

Van Imhoff, E. (1982). *Optimal economic growth and non-stable population*. Springer-Verlag, Berlin, Germany. 170

Wang, O. (2018). Banks, low interest rates, and monetary policy transmission. working paper, available at SSRN 3520134. 1, 4, 7, 9, 23, 26, 31

Wang, Y., Whited, T. M., Wu, Y., and Xiao, K. (2020). Bank market power and monetary policy transmission: Evidence from a structural estimation. Working Paper 27258, National Bureau of Economic Research. 7, 31, 41, 42

Werning, I. (2015). Incomplete markets and aggregate demand. Working Paper 21448, National Bureau of Economic Research. 252

Williamson, S. D. (2012). Liquidity, monetary policy, and the financial crisis: A new monetarist approach. *American Economic Review*, 102(6):2570–2605. 22

Williamson, S. D. (2019). Interest on reserves, interbank lending, and monetary policy. *Journal of Monetary Economics*, 101:14–30. 100

Wu, J. C. and Xia, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking*, 48(2-3):253–291. 10, 11

Yankov, V. (2014). In search of a risk-free asset. FEDS Working Paper No. 2014-108. 8