Disintermediating the Federal Funds Market

Tsz-Nga Wong
FRB Richmond

Mengbo Zhang
UCLA

This version: January 2022.
Check the latest version

Federal funds market, the interbank funding market implementing monetary policy, has been shrinking for more than 80% over a decade. We document and relate the decline to a new channel mediating the effects of unconventional monetary policy and bank regulation, the disintermediation channel. Observing that more than 80% of the Federal funds trades were once intermediated, when the policy spread over the interest rate on excess reserves (IOER) decreases, fewer banks intermediate in the Federal funds market, and if they do they intermediate less. In the data, the number of intermediating banks reduced from 600 to less than 100, and their purchases of Federal funds dropped by 90%. The disintermediation channel of the policy spread is significant and dominates other effects like expanded reserves and regulations. To explain this channel, we develop a continuous-time search-and-bargaining model of divisible assets and endogenous search intensity that includes many matching models as special cases. We solve the equilibrium in closed form, derive the dynamic distributions of trades and Federal fund rates, and the stopping times of entry and exit from the Federal funds market. In general, the equilibrium is constrained inefficient, as banks do not internalize the social surplus and intermediate *too often and too much*.

JEL Classification: G1, C78, D83, E44

Keywords: unconventional monetary policy, wholesale funding, regulation, search and bargaining, constrained inefficiency

Acknowledgments: We thank Fernando Alvarez, Andy Atkeson, Ricardo Lagos, Jeff Lacker, Guillaume Rocheteau, Pierre-Olivier Weill, Shengxing Zhang and participants of various seminars for comments and suggestion. The views expressed here are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

1 Introduction

Federal funds market, probably the most important market for monetary policy implementation, has been continuously declining since 2008, in spite of the fact that the American banks are larger than ever after the Great Recession. The market volume of Federal funds by the end of 2019 (before the pandemic) has dropped to less than 20% of its peak, while the total assets of banks have expanded by 4.5 times during the same period. Although it has become a concern of policymakers, this secular decline of the Federal funds market is still largely a puzzle (more on this later).

We propose a new mechanism with a search model to explain the puzzle: the disintermediation channel. Our new channel is motivated by an observation that Federal funds market was once heavily intermediated. A Federal funds trade is intermediated if the reserve-purchasing bank is also selling reserves on the same day, i.e., the bank borrows to lend.[1] Before 2008, typically more than 80% of the Federal funds were purchased by banks that intermediate. This feature is distinct from markets where the tradings of participants are usually one-sided. As illustrated in Figure 3, since the forth quarter of 2008, the decline in the Federal funds market is largely driven by the decline in the number of intermediating banks (fell from around 600 to 100), as well as their level of intermediated trades. We refer as to the *disintermediation* of the Federal funds market and explain with a search model. 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 & Lagos (2015b), banks always search for counterparties in the market, and they always trade to

[1] For readers not familiar with the jargons, Federal funds purchased are borrowing of reserves by buying the Federal funds today and selling back tomorrow. The price difference is the interest rate associated, i.e., the Federal funds rate. Similarly, Federal funds sold are lending of reserves.

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 & Krishnamurthy (2016), Bech & 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 & Stein (2012), Ennis (2018), Williamson (2019), Bigio & Sannikov (2021) and Bianchi & Bigio (ming).

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 & Monnet (2008), Bech & 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 & Lagos (2015b) and Afonso & Lagos (2015a). Our model endogenizes the time-varying search intensity to study the disintermediation trades. The related papers include Duffie et al. (2005), Lagos & Rocheteau (2009), Trejos & Wright (2016), Farboodi et al. (2017), Lagos & 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 & 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 & 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 describes the institutional background and the aggregate empirical facts that motivate our paper. Section 3 documents the empirical evidence on the disintermediation effect of unconventional monetary policies at the individual bank level. Section 4 presents the theoretical framework for our analysis. Section 5 provides a class of models that allows for closed-form solutions and comparative statics. Section 6 structurally estimates the analytical model and quantitatively decomposes the effects of unconventional monetary policies on Federal funds intermediation. The final section, 7, concludes the paper.

2 The Landscape of the Federal Funds Market

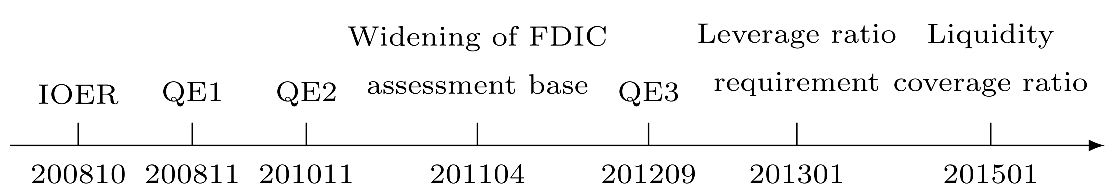
This section introduces the institutional features, the policy and regulatory environment and the aggregate trade dynamics in the Federal funds market to motivate our estimation and theoretical model in the following sections. We will focus on the change of the landscape of this market before and after 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.

2.1 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 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.

Figure 1: Timeline of unconventional monetary policy and regulation



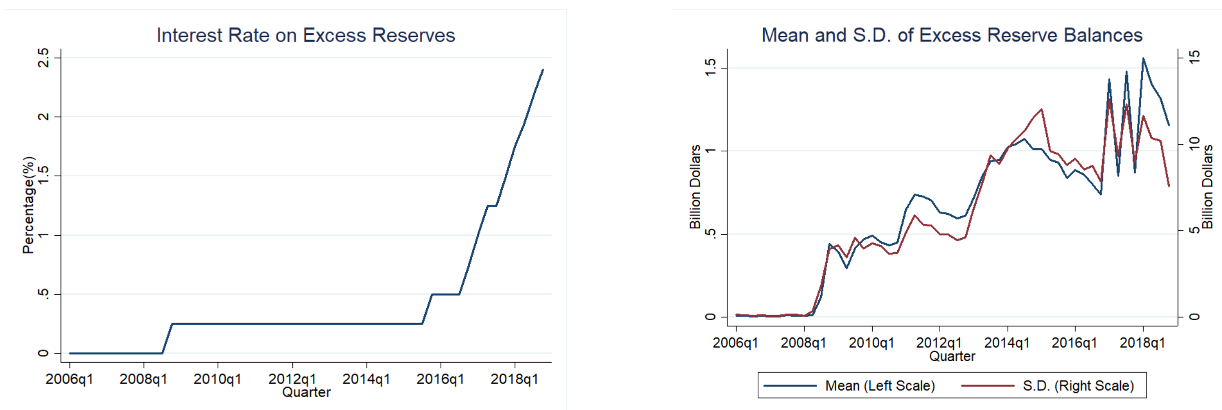
Notes: This figure plots the timeline of unconventional monetary policies and regulations since the Great Recession. The numbers on the timeline represents the date (year-month) when the policy or regulation is introduced.

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 including government-sponsored enterprises (GSEs) and some money market funds, is set at the bottom of the rage. Figure 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: 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.

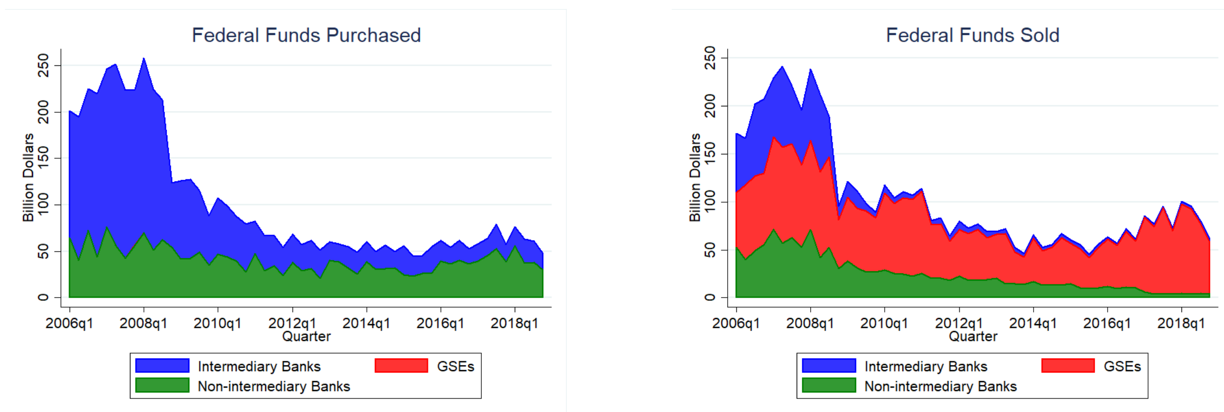
2.2 (Dis)intermediation in the Federal funds market

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 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 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.

Figure 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.

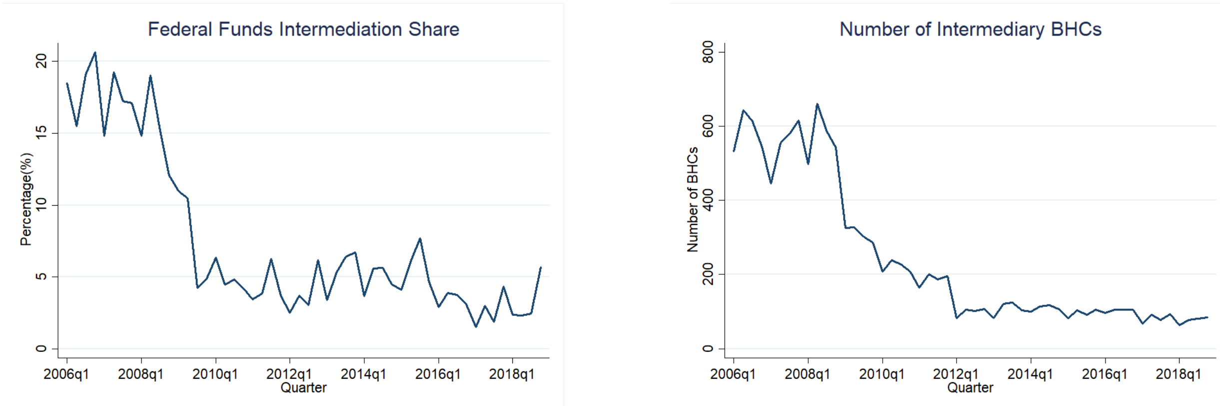
The decline of borrowing and lending by intermediaries imply the decline of Fed funds reallocation. We find that this decline occurs on both extensive and intensive margin. As plotted in Panel (a) of Figure 4, a significant share (more than 15%) of Fed funds volume is traded for intermediating purposes. However, since the financial crisis has declined by more than two thirds to less than 5%. Moreover, Panel (b) of Figure 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 1. All these changes closely relate to banks' incentive to trade Fed 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

Figure 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.

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

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 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 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.

3.1 Data

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).[2] 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 2018Q4.[3] 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.

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

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

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

This variable follows the definition of Fed funds reallocation in Afonso & 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 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 ($entity\_type$).

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);[4] (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.

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

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:

$$\begin{aligned} y_{i,t} \quad = \quad & fixed\_effects + \beta_0 exres\_assets_{i,t} + \beta_1 ioer_t + \beta_2 ioer_t \times exres\_assets_{i,t} \\ & + \beta_3 agg\_exres\_assets_t + \beta_4 agg\_exres\_assets_t \times exres\_assets_{i,t} \\ & + \beta_5 dw_t + \beta_6 dw_t \times exres\_assets_{i,t} + \gamma \cdot controls_{i,t} + \varepsilon_{i,t}, \end{aligned} \tag{1}$$

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 & Steinsson (2018).[5] 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 3, where we report three groups of estmation: column (1) and (2) reports the standard Probit estimation, Column (3) and (4) report the

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

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 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 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.

3.3 Effects on Net Borrowing of Fed 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:

$$\begin{aligned} ffnet\_assets_{i,t} \quad = \quad & \alpha_i + \eta_{yr(t)} + \beta_0 exres\_assets_{i,t} + \beta_1 ioer_t + \beta_2 ioer_t \times exres\_assets_{i,t} \\ & + \beta_3 agg\_exres\_assets_t + \beta_4 agg\_exres\_assets_t \times exres\_assets_{i,t} \\ & + \beta_5 dw_t + \beta_6 dw_t \times exres\_assets_{i,t} + \gamma \cdot controls_{i,t} + \varepsilon_{i,t}, \end{aligned} \tag{2}$$

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 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), 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 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.

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 & Lagos (2015b), but we endogenize the banks' search intensity.[6] 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

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

$T+\Delta$.[7] 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:

$$k_t = k_0 + \sum_{t_n \leq t} q_{t_n}, \tag{3}$$

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 ε_t and ε_t' are their search intensities.[8] 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

$$m\left[\varepsilon, \alpha\varepsilon' + (1-\alpha)\varepsilon''\right] = \alpha m\left(\varepsilon, \varepsilon'\right) + (1-\alpha) m\left(\varepsilon, \varepsilon''\right). \tag{4}$$

Define the search profile of all k-banks as $\boldsymbol{\varepsilon}_t = \{\varepsilon_t(k)\}_{k\in\mathbb{K}}$. 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 $m(\varepsilon, \varepsilon') = \lambda_0 + (\lambda - \lambda_0)\varepsilon\varepsilon'$.[9] 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

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

[8] 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

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

[9] The general form of the matching function is

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

See Appendix C.1 for derivations.

found are the same, which are proportional to both the bank's and the counterparty's search intensity.

Preferences. The individual bank's problem is given by

$$\max_{\boldsymbol{\varepsilon}} \mathbb{E}^{\boldsymbol{\varepsilon}} \left\{ \int_0^T e^{-rt} u(k_t)\, dt + e^{-rT} U(k_T) - \sum_{n=1,2,\ldots} \left[e^{-rt_n} \chi(\varepsilon_{t_n}, q_{t_n}) + e^{-r(T+\Delta)} R_{t_n} \right] \right\}, \text{ s.t. } (3). \tag{5}$$

The terms in the brackets of (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 (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 $(\boldsymbol{\varepsilon})$ to maximize the expected discounted payoff (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 & 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 continuation value of a bank holding k units of reserve balances at t.[10] 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:

$$\max_{\substack{q, R \in \mathbb{R} \\ k+q, k'-q \in \mathbb{K}}} B_t(k, q, R, \varepsilon)\, B_t\left(k', -q, -R, \varepsilon'\right). \tag{6}$$

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

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

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

$$V_t(k) = \mathbb{E}^{\boldsymbol{\varepsilon}} \left\{ \begin{array}{l} \int_0^{\min\{t_{+1}, T\}-t} e^{-r\tau} u(k)\, d\tau + 1_{t_{+1}>T} e^{-r(T-t)} U(k) \\ +1_{t_{+1} \leq T} e^{-r(t_{+1}-t)} \int \left\{ \begin{array}{l} V_{t_{+1}}\left[k + q_{t_{+1}}\left[k, k', \varepsilon_{t_{+1}}, \varepsilon_{t_{+1}}(k')\right]\right] \\ -\chi\left[\varepsilon_{t_{+1}}, q_{t_{+1}}\left[k, k', \varepsilon_{t_{+1}}, \varepsilon_{t_{+1}}(k')\right]\right] \\ -e^{-r(T+\Delta-t_{+1})} R_{t_{+1}}\left[k, k', \varepsilon_{t_{+1}}, \varepsilon_{t_{+1}}(k')\right] \end{array} \right\} \frac{m\left[\varepsilon_{t_{+1}}, \varepsilon_{t_{+1}}(k')\right]}{m\left(\varepsilon_{t_{+1}}, \bar{\varepsilon}_{t_{+1}}\right)} dF_{t_{+1}}(k') \end{array} \right\}, \tag{7}$$

where

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

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

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

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

$$B_t\left[k, q_t\left(k, k', \varepsilon, \varepsilon'\right), R_t\left(k, k', \varepsilon, \varepsilon'\right), \varepsilon\right] = B_t\left[k', -q_t\left(k, k', \varepsilon, \varepsilon'\right), -R_t\left(k, k', \varepsilon, \varepsilon'\right), \varepsilon'\right] = 0.5 S_t\left(k, k', \varepsilon, \varepsilon'\right).$$

Given the assumption on the cost function χ, the following lemma characterizes the property of the bargaining solution.[11]

Lemma 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 (7) can be recursively expressed as the solution the following Hamiltonian-

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

Jacob-Bellman (HJB) equation[12]

$$rV_t(k) = \dot{V}_t(k) + u(k) + \max_{\varepsilon_t \in [0,1]} \int \frac{1}{2} S_t\left[k, k', \varepsilon_t, \varepsilon_t(k')\right] m\left[\varepsilon_t, \varepsilon_t(k')\right] dF_t(k'), \text{ where } V_T(k) = U(k). \tag{8}$$

The initial value $V_0(k_0)$ equals (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)[13]

$$\dot{F}_t(k^w) = \left\{ \begin{array}{c} \int_{k>k^w} \int m[\varepsilon_t(k), \varepsilon_t(k')] 1\{k + q_t(k,k') \le k^w\} dF_t(k') dF_t(k) \\ -\int_{k\le k^w} \int m[\varepsilon_t(k), \varepsilon_t(k')] 1\{k + q_t(k,k') > k^w\} dF_t(k') dF_t(k) \end{array} \right\}, \text{ given } F_0(k^w). \tag{9}$$

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 (9) is the inflow rate to $I_-(k^w)$ post-trade from $I_+(k^w)$ pre-trade; the second line of (9) is the outflow rate from $I_-(k^w)$ pre-trade to $I_+(k^w)$ post-trade.

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 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 (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 (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 (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.[14] Second, due to the search complementarity (m is supermodular), the higher search intensities put by other banks the higher

[12] For readers not familiar with the HJB equation, we derive (8) in the online Appendix C.3. The discretized version of (8) without search cost or transaction cost is Proposition 1 of Afonso & Lagos (2015b).

[13] For readers not familiar with the KFE, we derive (9) in the online Appendix C.3. When k is discrete, $F_t(k)$ is probability mass function shown in Proposition 2 of Afonso & Lagos (2015b).

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

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

marginal propensity to match. Third, due to the cost shifting (S 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 (8), the equilibrium search profile is a fixed point function to the following functional:

$$\Gamma_t(\boldsymbol{\varepsilon}_t)(k) \equiv \arg\max_{\varepsilon\in\{0,1\}}\left\{\int S_t\left[k,k',\varepsilon,\varepsilon_t(k')\right]m\left[\varepsilon,\varepsilon_t(k')\right]dF_t(k')\right\}. \tag{10}$$

Denote the set of fixed points to Γ_t as $\Omega(S_t,F_t)\subseteq[0,1]^{\mathbb{K}}$, i.e., $\varepsilon_t(k)=\Gamma_t(\boldsymbol{\varepsilon}_t)(k)$ for all $\boldsymbol{\varepsilon}_t\in\Omega(S_t,F_t)$. 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 $\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}'\in\mathbb{L}\subseteq[0,1]^{\mathbb{K}}$, we have either $\boldsymbol{\varepsilon}\succeq_s'\boldsymbol{\varepsilon}$ or $\boldsymbol{\varepsilon}'\succeq_s\boldsymbol{\varepsilon}$. Suppose S_t satisfies the conditions for supermodularity in $(\varepsilon,\varepsilon')$ as in Lemma 1. The following proposition provides a sufficiently condition for complete lattice.

Proposition 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 $X(k,k',\varepsilon,\varepsilon')\geq1$ 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 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 *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 $t\in[0,T]$.*

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 & 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.

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 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.

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 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*

$$\mathbb{W} = \max\left\{\begin{array}{l} \int_0^T e^{-rt}\int u(k)\,dF_t^p(k)\,dt + e^{-rT}\int U(k)\,dF_T^p(k) \\ -\int_0^T\int\int e^{-rt}\chi\left[\varepsilon_t^p(k), q_t^p(k,k')\right] m\left[\varepsilon_t^p(k),\varepsilon_t^p(k')\right] dF_t^p(k')\,dF_t^p(k)\,dt \end{array}\right\} \tag{11}$$

subject to the law of motion of reserves

$$\dot{F}_t^p(k^w) = \left\{\begin{array}{l} \int_{k>k^w}\int m\left[\varepsilon_t^p(k),\varepsilon_t^p(k')\right]1\left\{k+q_t^p(k,k')\le k^w\right\}dF_t^p(k')\,dF_t^p(k) \\ -\int_{k\le k^w}\int m\left[\varepsilon_t^p(k),\varepsilon_t^p(k')\right]1\left\{k+q_t^p(k,k')> k^w\right\}dF_t^p(k')\,dF_t^p(k) \end{array}\right\}, \tag{12}$$

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

$$\begin{aligned} \mathcal{H}_t^p \equiv & \int u(k)\,dF_t^p(k) - \int\int \chi\left[\varepsilon_t^p(k), q_t^p(k,k')\right] m\left[\varepsilon_t^p(k),\varepsilon_t^p(k')\right] dF_t^p(k')\,dF_t^p(k) \\ & + \int\int m\left[\varepsilon_t^p(k),\varepsilon_t^p(k')\right]\left\{V_t^p\left[k+q_t^p(k,k')\right]-V_t^p(k)\right\}dF_t^p(k')\,dF_t^p(k) \\ & + \int\int \eta_t(k,k')\left[q_t^p(k,k')-q_t^p(k',k)\right]dF_t^p(k')\,dF_t^p(k), \end{aligned} \tag{13}$$

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[15]

$$\begin{aligned} rV_t^p(k) = & \dot{V}_t^p(k) + u(k) \\ & + \int\left\{\begin{array}{l} V_t^p\left[k+q_t^p(k,k')\right]+V_t^p\left[k'-q_t^p(k,k')\right]-V_t^p(k) \\ -V_t^p(k')-\chi\left[\varepsilon_t^p(k),q_t^p(k,k')\right]-\chi\left[\varepsilon_t^p(k'),-q_t^p(k,k')\right] \end{array}\right\} m\left[\varepsilon_t^p(k),\varepsilon_t^p(k')\right]dF_t^p(k'), \end{aligned} \tag{14}$$

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

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

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

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

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

where

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

Note that the equilibrium HJB (8) for $V_t(k)$ differs from the co-state HJB (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 ***(Inefficiency)*** *Equilibrium is not generically constrained optimal. Equilibrium is constrained optimal if* $\chi = 0$.

Afonso & Lagos (2015b) show that the welfare theorem holds when banks are homogeneous (beyond initial balances); Proposition 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 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.

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

$$da_t = \frac{\dot{\rho}_t^w}{1+\rho_t^w} a_t dt + d\delta_t, \tag{15}$$

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 (5), given the path of competitive Federal funds rates $\{\rho_t^w\}$, the bank problem is given by

$$\max_{\{\delta_t\}} \mathbb{E}\left\{\int_0^T e^{-rt}u\left(\frac{a_t}{1+\rho_t^w}\right)dt + e^{-rT}U\left(\frac{a_T}{1+\rho_T^w}\right) - e^{-r(T+\Delta)}\delta_T\right\}, \text{ s.t. (15).} \tag{16}$$

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

$$0 = \int \delta_t\left[(1+\rho_0^w)k\right]dF_0(k). \tag{17}$$

Proposition 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) $\delta_t(a) = (1+\rho_0^w)K - a$ *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.

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

$$\begin{aligned} u(k) &= -a_2k^2 + a_1k, \\ U(k) &= -A_2k^2 + A_1k. \end{aligned}$$

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

$$q_t\left(k, k', \varepsilon, \varepsilon'\right) = \left\{1 - \underbrace{\left[1 - \frac{V_t''(k) + V_t''(k')}{2\kappa\left(\varepsilon + \varepsilon'\right)}\right]^{-1}}_{\text{precaution-speed trade-off}}\right\} \underbrace{\frac{k' - k}{2}}_{\text{efficient bilateral trade}}, \tag{18}$$

$$1 + \rho_t\left(k, k', \varepsilon, \varepsilon'\right) = \underbrace{e^{r(T+\Delta-t)}}_{\text{time cost}} \left[\underbrace{\frac{V_t'(k) + V_t'(k')}{2}}_{\text{sharing marginal valuation}} + \underbrace{\kappa\frac{\varepsilon' - \varepsilon}{2} q_t\left(k, k', \varepsilon, \varepsilon'\right)}_{\text{speed premium (discount)}}\right]. \tag{19}$$

In Afonso & 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

$$\Gamma_t\left(\boldsymbol{\varepsilon}_t\right)(k) \equiv \arg\max_{\varepsilon\in[0,1]}\left\{\int\left[\frac{k'-k}{2}V_t''(k)\right]^2\underbrace{\frac{m\left(\varepsilon,\varepsilon_t\left(k'\right)\right)}{\kappa\left[\varepsilon+\varepsilon_t\left(k'\right)\right]}}_{\text{search efficiency}}\underbrace{\left[1-\frac{V_t''(k)+V_t''\left(k'\right)}{2\kappa\left[\varepsilon+\varepsilon_t\left(k'\right)\right]}\right]^{-1}}_{\text{precaution-speed trade-off}}dF_t\left(k'\right)\right\}, \tag{20}$$

Using Proposition 1, $\{\Omega\left(S_t,F_t\right),\succeq_s\}$ is a complete lattice since

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

and

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

Thus, the equilibrium search profile can be ranked.

Proposition 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., $\left|\Omega\left(S_t,F_t\right)\right|$, is greater than or equal to 1. When $\left|\Omega\left(S_t,F_t\right)\right|>1$, $\varepsilon_t(k)=1\ \forall k$ 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 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*

$$\begin{aligned}
\dot{H}_t &= rH_t - a_2 + \frac{1}{4}\frac{H_t^2}{\kappa\varepsilon_t + H_t}\left[(\lambda - \lambda_0)\varepsilon_t^2 + \lambda_0\right], \text{ where } H_T = A_2; && (21)\\
\dot{E}_t &= rE_t - a_1 + \frac{K}{2}\frac{H_t^2}{\kappa\varepsilon_t + H_t}\left[(\lambda - \lambda_0)\varepsilon_t^2 + \lambda_0\right], \text{ where } E_T = A_1; && (22)\\
\dot{D}_t &= rD_t - \frac{1}{4}\frac{H_t^2}{\kappa\varepsilon_t + H_t}\left[(\lambda - \lambda_0)\varepsilon_t^2 + \lambda_0\right]\int k'^2 dF_t(k'), \text{ where } D_T = 0. && (23)
\end{aligned}$$

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*

$$\begin{aligned}
q_t(k, k', \varepsilon, \varepsilon') &= \frac{H_t(k' - k)}{\kappa(\varepsilon + \varepsilon') + 2H_t}, && (24)\\
1 + \rho_t(k, k', \varepsilon, \varepsilon') &= e^{r(T+\Delta-t)}\left[E_t - H_t(k + k') - \frac{\kappa(\varepsilon - \varepsilon')}{2} q_t(k, k', \varepsilon, \varepsilon')\right]. && (25)
\end{aligned}$$

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 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

$$\varepsilon_t(k) = \begin{cases} 1, & \text{if } V_t''(k) \le -2\eta; \\ 0, & \text{otherwise.} \end{cases} \qquad (26)$$

Given Proposition 6, the following lemma solves the path of equilibrium search profile in the most liquid equilibrium.

Lemma 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

$$\varepsilon_t = \begin{cases} 1, & \text{if } t \geq \tau_1(\eta; A_2, T); \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

$$H_t = \begin{cases} \tau_1^{-1}(t; A_2, T), & \text{if } t \geq \tau_1(\eta; A_2, T); \\ J[t; \eta, \tau_1(\eta; A_2, T)], & \text{otherwise.} \end{cases} \tag{28}$$

(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

$$\varepsilon_t = \begin{cases} 0, & \text{if } t > \tau_2(\eta; A_2, T); \\ 1, & \text{otherwise.} \end{cases} \tag{29}$$

$$H_t = \begin{cases} J(t; A_2, T), & \text{if } t \geq \tau_2(\eta; A_2, T); \\ \tau_1^{-1}(t; \eta, \tau_2(\eta; A_2, T)), & \text{otherwise.} \end{cases} \tag{30}$$

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

$$\dot{F}_t(k) = m(\varepsilon_t, \varepsilon_t)\left[\int F_t\left[2\left(1 + \frac{\kappa\varepsilon_t}{H_t}\right)k - \left(1 + \frac{2\kappa\varepsilon_t}{H_t}\right)k'\right]dF_t(k') - F_t(k)\right], \tag{31}$$

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:

$$\dot{M}_{n,t} = m(\varepsilon_t, \varepsilon_t)\left[\sum_{i=0}^{n} C_n^i \frac{(H_t)^{n-i}(H_t + 2\kappa\varepsilon_t)^i}{2^n (H_t + \kappa\varepsilon_t)^n} M_{n-i,t} M_{i,t} - M_{n,t}\right], \tag{32}$$

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

$$M_{2,t} = K^2 + \left(M_{2,0} - K^2\right)\exp\left[-\int_0^t m(\varepsilon_z, \varepsilon_z)\frac{H_z(H_z + 2\kappa\varepsilon_z)}{2(H_z + \kappa\varepsilon_z)^2}dz\right]. \tag{33}$$

Thanks to Fourier transform, the model allows for an analytical expression for the paths of moments of the reserve distribution. In particular, equation (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.

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

$$\rho_t(k,q) \cong \underbrace{r(T+\Delta-t)}_{\text{time effect}} + \underbrace{\log V_t'(k)}_{\text{bank fixed effect}} - \underbrace{\frac{\theta_{V,t}(k)}{1-\omega_t}}_{\text{price impact}} \frac{q}{k}, \tag{34}$$

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:

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

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

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

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 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

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

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, $P_t^{q}(k)$ 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, $P^{q} \equiv \int P_0^{q}(k)\, dF(k)$ 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 $P^{\text{int}} = P^b + P^s - P^{q}$.

The laws of motion for the measures of trading banks, lending banks, borrowing banks, and intermediating banks are given by

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

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

$$P_t^{q}(k) = 1 - \exp\left[-\int_t^T m\left(\varepsilon_z, \varepsilon_z\right) dz\right]. \tag{35}$$

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:

$$Q_t(k) \equiv \mathbb{E} \sum_{t_i \in [t,T]} \left|q_{t_i}\left(k_{t_i}, k'\right)\right| \text{ s.t. } k_t = k. \tag{36}$$

The individual absolute trades follows

$$\dot{Q}_t(k) = -m\left(\varepsilon_t, \varepsilon_t\right)\left[\int_{k'} \left|q_t\left(k,k'\right)\right| dF_t\left(k'\right) + \int_{k'} Q_t\left(k + q_t\left(k,k'\right)\right) dF_t\left(k'\right) - Q_t(k)\right]. \tag{37}$$

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:

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

Thus the total trade volume is

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

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

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

The aggregate absolute net trade is defined as

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

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

$$L_t(k) = \left\{ 1 - \exp\left[- \int_t^T \frac{m(\varepsilon_z, \varepsilon_z) H_z}{2(\kappa \varepsilon_z + H_z)} dz \right] \right\} (K - k). \tag{38}$$

We can think of $L_t(k)$ as the net trade volume of bank k who contacts bank K at intensity $m(\varepsilon_t, \varepsilon_t)$. Thanks to the closed-form solution, we also derive the comparative statics of $L_t(k)$ on policy parameters in Section 5.3. Given the individual net trade volume, the aggregate volume of the absolute net trade is

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

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

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

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

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

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, 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 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

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

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

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 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*

	$T-\tau_1$	$\|q_t\|$	$L_0(k)$	$L_0'(k)$	$\rho_t(k,k')$
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

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

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

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 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 8 also produces the comparative statics of net Federal funds purchase that are consistent with the empirical evidence.

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 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*

$$\begin{aligned} & rV_t^p(k) \\ = \; & \dot{V}_t^p(k) + u(k) + \max_{\varepsilon\in[0,1]} \int_{k'} \max_{\substack{q\in\mathbb{R} \\ k+q,k'-q\in\mathbb{K}}} \left\{ \begin{array}{c} V_t^p(k+q) - V_t^p(k) - \chi(\varepsilon,q) \\ +V_t^p(k'-q) - V_t^p(k) - \chi(\varepsilon_t^p(k'),-q) \end{array} \right\} m\left(\varepsilon,\varepsilon_t^p(k')\right) dF_t^p(k') \end{aligned} \tag{39}$$

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

$$V_T^p(k) = U(k) \text{ for all } k\in\mathbb{K}, \tag{40}$$

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

$$\dot{F}_t^p(k^w) = \left\{ \begin{array}{c} \int_{k>k^w} \int m\left[\varepsilon_t^p(k),\varepsilon_t^p(k')\right] 1\{k+q_t^p(k,k')\le k^w\} dF_t^p(k') dF_t^p(k) \\ -\int_{k\le k^w} \int m\left[\varepsilon_t^p(k),\varepsilon_t^p(k')\right] 1\{k+q_t^p(k,k')> k^w\} dF_t^p(k') dF_t^p(k) \end{array} \right\}, \tag{41}$$

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

Note that the maximization problem in the planner's HJB (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 & 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 & 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

$$V_t^p(k) = -H_t^p k^2 + E_t^p k + D_t^p. \tag{42}$$

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

$$q_t^p\left(k,k'\right) = \frac{H_t^p(k'-k)}{2H_t^p + \kappa\left[\varepsilon_t^p(k)+\varepsilon_t^p(k')\right]}. \tag{43}$$

The bilateral surplus is

$$S_t^p\left(k,k'\right) = \frac{\left[H_t^p(k'-k)\right]^2}{2H_t^p + \kappa\left[\varepsilon_t^p(k)+\varepsilon_t^p(k')\right]}. \tag{44}$$

The optimal search intensity satisfies

$$\Gamma_t^p(\varepsilon_t^p)(k) \equiv \arg\max_{\varepsilon\in[0,1]} \int_{k'} S_t^p\left(k,k',\varepsilon,\varepsilon_t^p(k')\right) m\left(\varepsilon,\varepsilon_t^p(k')\right) dF_t^p\left(k'\right) \tag{45}$$

Therefore the HJB (39) simplifies to

$$rV_t^p(k) = \dot{V}_t^p(k) + u(k) + \frac{\left(H_t^p\right)^2}{2} \frac{(\lambda-\lambda_0)\left(\bar{\varepsilon}_t^p\right)^2 + \lambda_0}{\kappa\bar{\varepsilon}_t^p + H_t^p} \int_{k'} \left(k-k'\right)^2 dF_t^p\left(k'\right). \tag{46}$$

Matching coefficients yields

$$\dot{H}_t^p = rH_t^p - a_2 + \frac{\left(H_t^p\right)^2}{2} \frac{(\lambda-\lambda_0)\left(\bar{\varepsilon}_t^p\right)^2+\lambda_0}{\kappa\bar{\varepsilon}_t^p + H_t^p}, \text{ with } H_T^p = A_2. \tag{47}$$

$$\dot{E}_t^p = rE_t^p - a_1 + K\left(H_t^p\right)^2 \frac{(\lambda-\lambda_0)\left(\bar{\varepsilon}_t^p\right)^2+\lambda_0}{\kappa\bar{\varepsilon}_t^p + H_t^p}, \text{ with } E_T^p = A_1. \tag{48}$$

$$\dot{D}_t^p = rD_t^p - \frac{\left(H_t^p\right)^2}{2} \frac{(\lambda-\lambda_0)\left(\bar{\varepsilon}_t^p\right)^2+\lambda_0}{\kappa\bar{\varepsilon}_t^p + H_t^p} \int_{k'} \left(k'\right)^2 dF_t^p\left(k'\right), \text{ with } D_T^p = 0. \tag{49}$$

Note that the initial-value problem (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 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

$$\varepsilon_t^p = \begin{cases} 1, & \textit{if } t \geq \tau_1^p(\eta; A_2, T); \\ 0, & \textit{otherwise.} \end{cases} \tag{50}$$

$$H_t^p = \begin{cases} \left(\tau_1^p\right)^{-1}(t; A_2, T), & \textit{if } t \geq \tau_1^p(\eta; A_2, T); \\ J^p\left[t; \eta, \tau_1^p(\eta; A_2, T)\right], & \textit{otherwise.} \end{cases} \tag{51}$$

(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

$$\varepsilon_t^p = \begin{cases} 0, & \text{if } t > \tau_2^p(\eta; A, T); \\ 1, & \text{otherwise.} \end{cases} \tag{52}$$

$$H_t^p = \begin{cases} J^p(t; A_2, T), & \text{if } t \geq \tau_2^p(\eta; A, T); \\ (\tau_1^p)^{-1}(t; \eta, \tau_2^p(\eta; A, T)), & \text{otherwise.} \end{cases} \tag{53}$$

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

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 10 *For case (a-i) and (b-i) in Lemma 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 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 11 *(1) Suppose $A_2 < \eta^p$, $a_2 > \left(r + \frac{\lambda_0}{4}\right)\eta^p$, $\tau_2^p(\eta^p; A_2, T) > 0$, and T is sufficiently small. The comparative statics of the length of search, $\tau_2^p(\eta^p; A_2, T)$, the amount of Federal funds purchased, $q_t^p(k, k')$, net Federal funds purchase, $L_0^p(k)$ and its derivative $L_0^{p\prime}(k)$, and the bilateral Federal fund rates, $\rho_t^p(k, k')$, with respect to i^{ER}, i^{DW}, κ, λ_0, λ and K, are given by the following table*

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

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

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

and

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

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

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

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

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

and

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

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 6 describes the model setup and presents the quantitative analysis, while Appendix D provides the derivations for the closed-form solutions.

We also provide other extensions in the appendix. Appendix 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 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 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.

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.[16] 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 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. Specifically, we assume κ_0 and γ change over years in the following form:

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

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}.

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 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

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

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

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

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) 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 6. The simulated moments are listed in Table 7 and 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.

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 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.

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 & 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.

References

Acosta, M. & Saia, J. (2020). Estimating the effects of monetary policy via high frequency factors. Columbia University working paper.

Adda, J. & Cooper, R. (2003). *Dynamic economics: quantitative methods and applications*. MIT press.

Afonso, G., Armenter, R., & Lester, B. (2019). A model of the federal funds market: yesterday, today, and tomorrow. *Review of Economic Dynamics*, 33, 177–204.

Afonso, G., Kovner, A., & Schoar, A. (2011). Stressed, not frozen: The federal funds market in the financial crisis. *The Journal of Finance*, 66(4), 1109–1139.

Afonso, G. & 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.

Afonso, G. & Lagos, R. (2015b). Trade dynamics in the market for federal funds. *Econometrica*, 83(1), 263–313.

Bech, M. & Keister, T. (2017). Liquidity regulation and the implementation of monetary policy. *Journal of Monetary Economics*, 92, 64–77.

Bech, M. & Monnet, C. (2016). A search-based model of the interbank money market and monetary policy implementation. *Journal of Economic Theory*, 164, 32–67.

Bech, M. L. & Atalay, E. (2010). The topology of the federal funds market. *Physica A: Statistical Mechanics and its Applications*, 389(22), 5223–5246.

Benhabib, J. & Farmer, R. E. (1994). Indeterminacy and increasing returns. *Journal of Economic Theory*, 63(1), 19–41.

Berentsen, A. & Monnet, C. (2008). Monetary policy in a channel system. *Journal of Monetary Economics*, 55(6), 1067–1080.

Bianchi, J. & Bigio, S. (forthcoming). Banks, liquidity management and monetary policy. *Econometrica*.

Bigio, S. & Sannikov, Y. (2021). *A Model of Credit, Money, Interest, and Prices*. Working Paper 28540, National Bureau of Economic Research.

Bracewell, R. N. (2000). *The Fourier transform and its applications*. McGraw-Hill New York.

Chang, B. & Zhang, S. (2018). Endogenous market making and network formation. Available at SSRN 2600242.

Chiu, J., Eisenschmidt, J., & Monnet, C. (2020). Relationships in the interbank market. *Review of Economic Dynamics*, 35, 170–191.

Duffie, D., Gârleanu, N., & Pedersen, L. H. (2005). Over-the-counter markets. *Econometrica*, 73(6), 1815–1847.

Duffie, D. & 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* (pp. 1815–1847).

Ennis, H. M. (2018). A simple general equilibrium model of large excess reserves. *Journal of Monetary Economics*, 98, 50–65.

Farboodi, M., Jarosch, G., & Shimer, R. (2017). *The emergence of market structure*. Working Paper 23234, National Bureau of Economic Research.

Gofman, M. (2017). Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *Journal of Financial Economics*, 124(1), 113–146.

Hamilton, J. D. (1996). The daily market for federal funds. *Journal of Political Economy*, 104(1), 26–56.

Hugonnier, J., Lester, B., & Weill, P.-O. (2020). Frictional intermediation in over-the-counter markets. *The Review of Economic Studies*, 87(3), 1432–1469.

Ihrig, J. E., Vojtech, C. M., & Weinbach, G. C. (2019). How have banks been managing the composition of high-quality liquid assets? *Review*, 101(3), 177–201.

Kashyap, A. K. & Stein, J. C. (2012). The optimal conduct of monetary policy with interest on reserves. *American Economic Journal: Macroeconomics*, 4(1), 266–82.

Keating, T. & Macchiavelli, M. (2017). Interest on reserves and arbitrage in post-crisis money markets. FEDS Working Paper No. 2017-124.

Lagos, R. & Rocheteau, G. (2007). Search in asset markets: Market structure, liquidity, and welfare. *American Economic Review*, 97(2), 198–202.

Lagos, R. & Rocheteau, G. (2009). Liquidity in asset markets with search frictions. *Econometrica*, 77(2), 403–426.

Lagos, R. & Zhang, S. (2019). A monetary model of bilateral over-the-counter markets. *Review of Economic Dynamics*, 33, 205–227.

Liu, S. (2020). Dealers' search intensity in us corporate bond markets. Available at SSRN 3644132.

Milgrom, P. & Shannon, C. (1994). Monotone comparative statics. *Econometrica*, 62(1), 157–180.

Nakamura, E. & Steinsson, J. (2018). High-frequency identification of monetary non-neutrality: the information effect. *The Quarterly Journal of Economics*, 133(3), 1283–1330.

Poole, W. (1968). Commercial bank reserve management in a stochastic model: implications for monetary policy. *The Journal of finance*, 23(5), 769–791.

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.

Trejos, A. & Wright, R. (2016). Search-based models of money and finance: An integrated approach. *Journal of Economic Theory*, 164, 10–31.

Üslü, S. (2019). Pricing and liquidity in decentralized asset markets. *Econometrica*, 87(6), 2079–2140.

van Imhoff, E. (1982). *Optimal economic growth and non-stable population*. Springer-Verlag, Berlin, Germany.

Williamson, S. D. (2019). Interest on reserves, interbank lending, and monetary policy. *Journal of Monetary Economics*, 101, 14–30.

Appendices

A 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.

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).[17]
- **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.[18]
- **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).[19]
- **10Q and 10K**. This is the source of government sponsored enterprises (GSE) data. These forms are available for download at the Security Examination Commission (SEC).[20] 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.[21]
- **Time series of the economy**. It is available for download at the Federal Reserve Bank of St Louis (FRED).[22]

[17] https://cdr.ffiec.gov/public/
[18] https://www.chicagofed.org/banking/financial-institution-reports/bhc-data
[19] https://www.ffiec.gov/npw/FinancialReport/DataDownload
[20] https://www.sec.gov/edgar/searchedgar/companysearch.html
[21] https://www.federalreserve.gov/releases/h41/
[22] https://fred.stlouisfed.org/

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.

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

$$\textit{Excess Reserves}_{it} = \textit{Total Reserves}_{it} \text{ - } \left\{ \sum\nolimits_j \textit{Required Reserves}_{jt} \text{ - } \textit{Vault Cash}_{it} \right\}_+ .$$

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 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.[23] 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

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

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 $\textit{Required Reserves}_{jt}$.

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

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

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

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 $\textit{Reallocated Funds}_{it}$:

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

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.

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)

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)

B Tables

B.1 Summary statistics

Table 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.

B.2 Regression results

Table 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×Ind. ex res		-0.480*		-0.625		-0.634
		(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 (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 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×Ind. ex res		-0.000		0.001		-0.001
		(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 (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 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×Ind. ex res		-0.071***	-0.071***		-0.059***	-0.059***
		(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×Crisis FE	Y	Y	Y	Y	Y	Y
Bank×Post-Crisis FE	Y	Y	Y	Y	Y	Y
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
Relative OLS bias>10% (p-value)				[0.000]	[0.000]	[0.000]
Relative OLS bias>30% (p-value)				[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). 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. Standard errors clustered by banks are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

B.3 Tables in Quantitative Analysis

Table 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 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 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 9: Counterfactual analysis

Moments in 2018	Target	Simulation	Counterfactual analysis (1) IOER	(2) Agg ex res	(3) 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.

C Proofs and Derivations

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

For any $\varepsilon, \varepsilon' \in [0,1]$, equation (4) implies that

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

By symmetry we have

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

Thus we can get

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

C.2 Proof of Lemma 1

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

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

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

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

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),

$$q_t\left(k, k', \varepsilon, \varepsilon'\right) = \arg\max_q \left\{V_t(k+q) + V_t\left(k'-q\right) - \chi(\varepsilon, q) - \chi\left(\varepsilon', q\right)\right\} \tag{C.1}$$

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

$$\frac{\partial S_t(k, k', \varepsilon, \varepsilon')}{\partial \varepsilon} = -\chi_{\varepsilon}\left(\varepsilon, q_t\left(k, k', \varepsilon, \varepsilon'\right)\right),$$

and

$$\frac{\partial^2 S_t(k, k', \varepsilon, \varepsilon')}{\partial \varepsilon \partial \varepsilon'} = -\chi_{\varepsilon q}\left(\varepsilon, q_t\left(k, k', \varepsilon, \varepsilon'\right)\right) \frac{\partial q_t(k, k', \varepsilon, \varepsilon')}{\partial \varepsilon'} > 0,$$

where we apply $\chi_{\varepsilon q} > 0$ and $\frac{\partial q_t(k, k', \varepsilon, \varepsilon')}{\partial \varepsilon'} < 0$.

(ii). [$S_t(k, k, \varepsilon, \varepsilon') = 0$] If $V_t(k)$ is concave, then

$$\begin{aligned}
S_t\left(k, k, \varepsilon, \varepsilon'\right) &\equiv \max_q \left\{V_t(k+q) + V_t(k-q) - V_t(k) - V_t(k) - \chi(\varepsilon, q) - \chi\left(\varepsilon', -q\right)\right\} \\
&\leq \max_q \left\{2V_t(k) - V_t(k) - V_t(k) - \chi(\varepsilon, q) - \chi\left(\varepsilon', -q\right)\right\} \\
&= \max_q \left\{-\chi(\varepsilon, q) - \chi\left(\varepsilon', q\right)\right\} = 0.
\end{aligned}$$

[Monotonicity of $S_t(k,k',\varepsilon,\varepsilon')$ and $q_t(k,k',\varepsilon,\varepsilon')$ in k] Note that for a concave functio $f(x)$, where x is a scalar, we must have that for any $x' > x$ and $\Delta > 0$,

$$\begin{aligned} f(x) + f(x') &= f\left(\frac{x'-x+\Delta}{x'-x+2\Delta}(x-\Delta) + \frac{\Delta}{x'-x+2\Delta}(x'+\Delta)\right) \\ &\quad + f\left(\frac{\Delta}{x'-x+2\Delta}(x-\Delta) + \frac{x'-x+\Delta}{x'-x+2\Delta}(x'-\Delta)\right) \\ &\geq \frac{x'-x+\Delta}{x'-x+2\Delta} f(x-\Delta) + \frac{\Delta}{x'-x+2\Delta} f(x'+\Delta) \\ &\quad + \frac{\Delta}{x'-x+2\Delta} f(x-\Delta) + \frac{x'-x+\Delta}{x'-x+2\Delta} f(x'+\Delta) \\ &= f(x-\Delta) + f(x'+\Delta). \end{aligned} \tag{C.2}$$

Thus for any $k' > k$ and $q < 0$:

$$\begin{aligned} & V_t(k+q) + V_t(k'-q) - \chi(\varepsilon,q) - \chi(\varepsilon',-q) \\ < \; & V_t(k) + V_t(k') - \chi(\varepsilon,0) - \chi(\varepsilon',0), \end{aligned}$$

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 (C.2) 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 & 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.**

C.3 Derivation of HJB (8) and KFE (9)

By the property of Poisson process, the equation (7) for value function $V_t(k)$ can be rewritten as

$$\begin{aligned} & V_t(k) \\ = & \max_{\{\varepsilon_z\}_{z\in[t,T]}\in[0,1]^{[t,T]}} \left\{ \begin{array}{l} \int_t^T e^{-\int_t^z [r+m(\varepsilon_s,\bar{\varepsilon}_s)]ds} \left\{ \begin{array}{l} u(k) + \int_{k'} \left\{ \begin{array}{l} V_z[k + q_z(k, k', \varepsilon_z, \varepsilon_z(k'))] \\ -\chi[\varepsilon_z, q_z(k, k', \varepsilon_z, \varepsilon_z(k'))] \\ -e^{-r(T+\Delta-z)} R_z(k, k', \varepsilon_z, \varepsilon_z(k')) \end{array} \right\} \\ \times m(\varepsilon_z, \varepsilon_z(k')) dF_z(k') \end{array} \right\} dz \\ + e^{-\int_t^T [r+m(\varepsilon_s,\bar{\varepsilon}_s)]ds} U(k) \end{array} \right\} \end{aligned}$$

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 $e^{-r(T+\Delta-z)} R_z(k, k', \varepsilon_z, \varepsilon_z(k'))$, 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:

$$\begin{aligned} & (\mathcal{M}w)(k, t) \\ = & \max_{\{\varepsilon_z\}_{z\in[t,T]}\in[0,1]^{[t,T]}} \left\{ \begin{array}{l} \int_t^T e^{-\int_t^z [r+m(\varepsilon_s,\bar{\varepsilon}_s)]ds} \left\{ \begin{array}{l} u(k) + \int_{k'} \left\{ \begin{array}{l} w[k + b_z(k, k', \varepsilon_z, \varepsilon_z(k')), z] \\ -\chi[\varepsilon_z, b_z(k, k', \varepsilon_z, \varepsilon_z(k'))] \\ -e^{-r(T+\Delta-z)} Y_z(k, k', \varepsilon_z, \varepsilon_z(k')) \end{array} \right\} \\ \times m(\varepsilon_z, \varepsilon_z(k')) dF_z(k') \end{array} \right\} dz \\ + e^{-\int_t^T [r+m(\varepsilon_s,\bar{\varepsilon}_s)]ds} U(k) \end{array} \right\} \end{aligned}$$

where

$$b_t\left(k, k', \varepsilon, \varepsilon\prime\right) \in \arg\max_b \left\{ \begin{array}{c} w(k + b, t) - w(k, t) - \chi(\varepsilon, b) \\ +w(k' - b, t) - w(k', t) - \chi(\varepsilon', -b) \end{array} \right\}$$

and

$$e^{-r(T+\Delta-t)} Y_t\left(k, k', \varepsilon, \varepsilon'\right) = \frac{1}{2} \left\{ \begin{array}{c} w(k + b_t(k, k', \varepsilon, \varepsilon'), t) - w(k, t) - \chi(\varepsilon, b_t(k, k', \varepsilon, \varepsilon')) \\ +w(k', t) - w(k' - b_t(k, k', \varepsilon, \varepsilon'), t) + \chi(\varepsilon', -b_t(k, k', \varepsilon, \varepsilon')) \end{array} \right\}.$$

It is clear that the solution $V_t(k)$ to the HJB (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

$$
\begin{aligned}
& e^{-\int_t^{t+dt}[r+m(\varepsilon_t(k),\bar{\varepsilon}_t)]ds}\left\{\begin{array}{l} u(k)+\int_{k'}\left\{\begin{array}{l} w\left[k+b_t\left(k,k',\varepsilon_t(k),\varepsilon_t(k')\right),t\right] \\ -\chi\left[\varepsilon_t(k),b_t\left(k,k',\varepsilon_t(k),\varepsilon_t(k')\right)\right] \\ -e^{-r(T+\Delta-t)}Y_t\left(k,k',\varepsilon_t(k),\varepsilon_t(k')\right) \end{array}\right\} \\ \times m\left(\varepsilon_t(k),\varepsilon_t(k')\right)dF_t(k') \end{array}\right\}dt \\
& +e^{-\int_t^{t+dt}[r+m(\varepsilon_t(k),\bar{\varepsilon}_t)]ds}w(k,t-dt) \\
= \; & (1-rdt)\,w(k,t-dt)+o(|dt|)+\left\{\begin{array}{l} u(k)+\int_{k'}\left\{\begin{array}{l} w\left[k+b_t\left(k,k',\varepsilon_t(k),\varepsilon_t(k')\right),t\right] \\ -w(k,t-dt) \\ -\chi\left[\varepsilon_t,b_t\left(k,k',\varepsilon_t(k),\varepsilon_t(k')\right)\right] \\ -e^{-r(T+\Delta-t)}Y_t\left(k,k',\varepsilon_t(k),\varepsilon_t(k')\right) \end{array}\right\} \\ \times m\left(\varepsilon_t(k),\varepsilon_t(k')\right)dF_t(k') \end{array}\right\}dt.
\end{aligned}
$$

Thus the maximizer of $\varepsilon_t(k)$ to the above equation when $dt \to 0$ is given by

$$
\varepsilon_t(k) \in \arg\max_{\varepsilon\in[0,1]}\left\{\begin{array}{l} \int_{k'}\frac{1}{2}\left[\begin{array}{l} w\left[k+b_t\left(k,k',\varepsilon,\varepsilon_t(k')\right),t\right]-w(k,t)-\chi\left[\varepsilon_t,b_t\left(k,k',\varepsilon,\varepsilon_t(k')\right)\right] \\ +w\left[k'-b_t\left(k,k',\varepsilon,\varepsilon_t(k')\right),t\right]-w(k',t)-\chi\left[\varepsilon_t(k'),-b_t\left(k,k',\varepsilon,\varepsilon_t(k')\right)\right] \end{array}\right] \\ \times m\left(\varepsilon_t(k),\varepsilon_t(k')\right)dF_t(k') \end{array}\right\}
$$

where we plug in the solution to $e^{-r(T+\Delta-t)}Y_t(k,k',\varepsilon,\varepsilon_t(k'))$. This gives the HJB (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

$$
\begin{aligned}
F_{t+\Delta}(k^w) \; = \; & \left[1-\Delta\cdot m\left(\varepsilon_t(k),\bar{\varepsilon}_t\right)\right]F_t(k^w) \\
& +\int_{k\le k^w}\int_{k'}\Delta\cdot m\left(\varepsilon_t(k),\varepsilon_t(k')\right)1\left\{k+q_t(k,k')\le k^w\right\}dF_t(k')\,dF_t(k) \\
& +\int_{k>k^w}\int_{k'}\Delta\cdot m\left(\varepsilon_t(k),\varepsilon_t(k')\right)1\left\{k+q_t(k,k')\le k^w\right\}dF_t(k')\,dF_t(k).
\end{aligned}
$$

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. These three groups of banks constitute the mass of banks with reserves no more than k^w at $t+\Delta$. By rearranging terms, we can get

$$
\begin{aligned}
\frac{F_{t+\Delta}(k^w)-F_t(k^w)}{\Delta} \; = \; & -m\left(\varepsilon_t(k),\bar{\varepsilon}_t\right)F_t(k^w) \\
& +\int_{k\le k^w}\int_{k'}m\left(\varepsilon_t(k),\varepsilon_t(k')\right)1\left\{k+q_t(k,k')\le k^w\right\}dF_t(k')\,dF_t(k) \\
& +\int_{k>k^w}\int_{k'}m\left(\varepsilon_t(k),\varepsilon_t(k')\right)1\left\{k+q_t(k,k')\le k^w\right\}dF_t(k')\,dF_t(k) \\
= \; & -\int_{k\le k^w}\int_{k'}m\left(\varepsilon_t(k),\varepsilon_t(k')\right)1\left\{k+q_t(k,k')> k^w\right\}dF_t(k')\,dF_t(k) \\
& +\int_{k>k^w}\int_{k'}m\left(\varepsilon_t(k),\varepsilon_t(k')\right)1\left\{k+q_t(k,k')\le k^w\right\}dF_t(k')\,dF_t(k),
\end{aligned}
$$

where in the second equality 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 (9).

C.4 Proof of Proposition 1

Proof. To prove $\{\Omega\left(S_t,F_t\right),\succeq_s\}$ 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 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.**

C.5 Derivation of Equation (14)

Following Üslü (2019), the planner's current-value Hamiltonian can be written as

$$\begin{aligned}
\mathcal{H}_t^p \equiv\;& \int u\left(k\right)dF_t^p\left(k\right)-\int\int\chi\left[\varepsilon_t^p\left(k\right),q_t^p\left(k,k'\right)\right]m\left[\varepsilon_t^p\left(k\right),\varepsilon_t^p\left(k'\right)\right]dF_t^p\left(k'\right)dF_t^p\left(k\right) \quad \text{(C.3)}\\
&+\int\int m\left[\varepsilon_t^p\left(k\right),\varepsilon_t^p\left(k'\right)\right]\left\{V_t^p\left[k+q_t^p\left(k,k'\right)\right]-V_t^p\left(k\right)\right\}dF_t^p\left(k'\right)dF_t^p\left(k\right)\\
&+\int\int\eta_t\left(k,k'\right)\left[q_t^p\left(k,k'\right)+q_t^p\left(k',k\right)\right]dF_t^p\left(k'\right)dF_t^p\left(k\right).
\end{aligned}$$

First-order conditions. First, take any optimal q_t^e and

$$\begin{aligned}\hat{q}_t\left(k,k'\right) &= q_t^e\left(k,k'\right)+\alpha_q\mathbf{1}\left\{V_t^e\left(k\right)>V_t^e\left(k'\right)\right\}-\alpha_q\mathbf{1}\left\{V_t^e\left(k\right)<V_t^e\left(k'\right)\right\}\\ &= q_t^e\left(k,k'\right)+\alpha_q\Delta_t\left(k,k'\right),\end{aligned}$$

where α_q is an arbitrary scalar. Second, take any optimal $\varepsilon_t^e\left(k\right)$, an arbitrary admissible deviation $\delta_t\left(k\right)$ and a scalar α_ε, let

$$\hat{\varepsilon}_t\left(k\right)=\varepsilon_t^e\left(k\right)+\alpha_\varepsilon\cdot\delta_t\left(k\right).$$

For small α_q and α_ε, we obtain up to second-order terms:

$$\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}\chi_1\left[\varepsilon_t^e\left(k\right),q_t^e\left(k,k'\right)\right]m\left[\varepsilon_t^e\left(k\right),\varepsilon_t^e\left(k'\right)\right]\delta_t\left(k\right)\\ +\chi\left[\varepsilon_t^e\left(k\right),q_t^e\left(k,k'\right)\right]m_1\left[\varepsilon_t^e\left(k\right),\varepsilon_t^e\left(k'\right)\right]\delta_t\left(k\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\left(k'\right)\right]\delta_t\left(k'\right)\end{array}\right\}dF_t^p\left(k'\right)dF_t^p\left(k\right)\\
&+\alpha_\varepsilon\int\int\left\{\begin{array}{c}m_1\left[\varepsilon_t^e\left(k\right),\varepsilon_t^e\left(k'\right)\right]\delta_t\left(k\right)\\ +m_2\left[\varepsilon_t^e\left(k\right),\varepsilon_t^e\left(k'\right)\right]\delta_t\left(k'\right)\end{array}\right\}\left\{V_t^p\left[k+q_t^e\left(k,k'\right)\right]-V_t^p\left(k\right)\right\}dF_t^p\left(k'\right)dF_t^p\left(k\right)\\
&-\alpha_q\int\int\chi_2\left[\varepsilon_t^e\left(k\right),q_t^e\left(k,k'\right)\right]m\left[\varepsilon_t^e\left(k\right),\varepsilon_t^e\left(k'\right)\right]\Delta_t\left(k,k'\right)dF_t^p\left(k'\right)dF_t^p\left(k\right)\\
&+\alpha_q\int\int m\left[\varepsilon_t^e\left(k\right),\varepsilon_t^e\left(k'\right)\right]V_t^{p\prime}\left[k+q_t^e\left(k,k'\right)\right]\Delta_t\left(k,k'\right)dF_t^p\left(k'\right)dF_t^p\left(k\right)\\
&+\alpha_q\int\int\eta_t\left(k,k'\right)\left[\Delta_t\left(k,k'\right)+\Delta_t\left(k',k\right)\right]dF_t^p\left(k'\right)dF_t^p\left(k\right)\\
=\;& \alpha_\varepsilon\int\int\left\{\begin{array}{c}m_1\left[\varepsilon_t^e\left(k\right),\varepsilon_t^e\left(k'\right)\right]\left\{V_t^p\left[k+q_t^e\left(k,k'\right)\right]-V_t^p\left(k\right)\right\}\\ +m_2\left[\varepsilon_t^e\left(k'\right),\varepsilon_t^e\left(k\right)\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\left(k\right),q_t^e\left(k,k'\right)\right]m\left[\varepsilon_t^e\left(k\right),\varepsilon_t^e\left(k'\right)\right]\\ -\chi\left[\varepsilon_t^e\left(k\right),q_t^e\left(k,k'\right)\right]m_1\left[\varepsilon_t^e\left(k\right),\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\left(k\right)\right]\end{array}\right\}\delta_t\left(k\right)dF_t^p\left(k'\right)dF_t^p\left(k\right)\\
&+\frac{\alpha_q}{2}\int\int m\left[\varepsilon_t^e\left(k\right),\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)dF_t^p\left(k'\right)dF_t^p\left(k\right)\\
&+\frac{\alpha_q}{2}\int\int m\left[\varepsilon_t^e\left(k\right),\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)dF_t^p\left(k'\right)dF_t^p\left(k\right)\\
=\;& \alpha_\varepsilon\int\int\left\{\begin{array}{c}m_1\left[\varepsilon_t^e\left(k\right),\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\}\\ +m_1\left[\varepsilon_t^e\left(k\right),\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\left(k\right),q_t^p\left(k,k'\right)\right]m\left[\varepsilon_t^e\left(k\right),\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\left(k\right),\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\left(k\right),\varepsilon_t^e\left(k'\right)\right]\end{array}\right\}\delta_t\left(k\right)dF_t^p\left(k'\right)dF_t^p\left(k\right)\\
&+\frac{\alpha_q}{2}\int\int m\left[\varepsilon_t^e\left(k\right),\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\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]+\chi_2\left[\varepsilon_t^e\left(k'\right),-q_t^e\left(k,k'\right)\right]\end{array}\right\}\Delta_t\left(k,k'\right)dF_t^p\left(k'\right)dF_t^p\left(k\right),
\end{aligned}$$

where we apply $\Delta_t\left(k,k'\right)+\Delta_t\left(k',k\right)=0$ in the second and third equality and $q_t^e\left(k,k'\right)+q_t^e\left(k',k\right)=0$ in the third equality.

If $\left\{\varepsilon_t^e,q_t^e\right\}$ is optimal, this 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

$$V_t^{p\prime}\left[k+q_t^e\left(k,k'\right)\right]-V_t^{p\prime}\left[k'-q_t^e\left(k,k'\right)\right]-\chi_2\left[\varepsilon_t^e\left(k\right),q_t^e\left(k,k'\right)\right]+\chi_2\left[\varepsilon_t^e\left(k'\right),-q_t^e\left(k,k'\right)\right]=0.$$

In other words, $q_t^p(k,k')$ is the solution to

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

Moreover, for the FOC of ε_t^e, since $\delta_t(k)$ is an arbitrary admissible deviation, we must have

$$\begin{array}{l} m_1\left[\varepsilon_t^e\left(k\right),\varepsilon_t^e\left(k'\right)\right]\left\{\begin{array}{c} V_t^p\left[k+q_t^p\left(k,k'\right)\right]-V_t^p\left(k\right)-\chi\left[\varepsilon_t^e\left(k\right),q_t^p\left(k,k'\right)\right] \\ +V_t^p\left[k'-q_t^p\left(k,k'\right)\right]-V_t^p\left(k'\right)-\chi\left[\varepsilon_t^e\left(k'\right),-q_t^p\left(k,k'\right)\right] \end{array}\right\} \\ -\chi_1\left[\varepsilon_t^e\left(k\right),q_t^p\left(k,k'\right)\right]m\left[\varepsilon_t^e\left(k\right),\varepsilon_t^e\left(k'\right)\right] \\ \left\{\begin{array}{ll} \le 0, & \text{if } \varepsilon_t^e\left(k\right)=0, \\ =0, & \text{if } \varepsilon_t^e\left(k\right)\in\left(0,1\right), \\ \ge 0, & \text{if } \varepsilon_t^e\left(k\right)=1. \end{array}\right. , \end{array}$$

Thus the constrained efficiency solution of ε_t^p must satisfy

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

where

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

C.6 Proof of Proposition 3

Proof. Denote v_t^w as the co-state to a_t, the Hamiltonian is thus given by

$$\mathcal{H}_t^w\equiv u\left(\frac{a_t}{1+\rho_t^w}\right)-e^{-r(T+\Delta-t)}d\delta_t+v_t^w\left(\frac{\dot\rho_t^w}{1+\rho_t^w}a_t+d\delta_t\right). \tag{C.4}$$

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.

$$\dot v_t^w=rv_t^w-\frac{1}{1+\rho_t^w}u'\left(\frac{a_t}{1+\rho_t^w}\right)-v_t^w\frac{\dot\rho_t^w}{1+\rho_t^w}. \tag{C.5}$$

The first order condition with respect to $d\delta_t$ is

$$v_t^w=e^{-r(T+\Delta-t)}. \tag{C.6}$$

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, C.5, 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 (C.6) to the evolution of costate, (C.5), we have

$$\dot\rho_t^w=-e^{r(T+\Delta-t)}u'\left(K\right).$$

The solution to the above ODE is

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

Notice that at T the bank problem is

$$\max_{q_T} \left\{ U\left(k+q_T\right) - e^{-r\Delta}\left(1+\rho_T^w\right) q_T \right\}.$$

To yield $k+q_T = K$, we have

$$\rho_T^w = e^{r\Delta} U'\left(K\right) - 1.$$

Q.E.D.

C.7 Proof of Proposition 4

Proof. [$\varepsilon_t\left(k\right)=0$ is always an equilibrium] For any k and ε, if all the other banks choose zero search intensity, then

$$\frac{\partial \int S_t\left(k,k',\varepsilon,0\right) m\left(\varepsilon,0\right) dF_t\left(k'\right)}{\partial \varepsilon} = \int \left[\frac{k'-k}{2} V_t''\left(k\right)\right]^2 \frac{-\kappa\lambda_0}{\left[\kappa\left(\varepsilon\right) - \frac{1}{2}\left(V_t''\left(k\right)+V_t''\left(k'\right)\right)\right]^2} dF_t\left(k'\right) < 0.$$

This implies that the bank k's optimal response is $\varepsilon = 0$. Thus $\varepsilon_t\left(k\right) = 0$ $\forall k$ 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 $\varepsilon_t\left(k\right)=1$ $\forall k$ to be an equilibrium. Suppose all the other banks choose search intensity $\varepsilon_t = 1$. Then for any k and ε, we have

$$\begin{aligned}
& \frac{\partial \int S_t\left(k,k',\varepsilon,1\right) m\left(\varepsilon,1\right) dF_t\left(k'\right)}{\partial \varepsilon} \\
= & \int \left[\frac{k'-k}{2} V_t''\left(k\right)\right]^2 \frac{\left(\lambda-\lambda_0\right)\kappa - \frac{\lambda-\lambda_0}{2}\left[V_t''\left(k\right)+V_t''\left(k'\right)\right] - \kappa\lambda_0}{\left[\kappa\left(\varepsilon+1\right) - \frac{1}{2}\left(V_t''\left(k\right)+V_t''\left(k'\right)\right)\right]^2} dF_t\left(k'\right) \\
= & \frac{\left(\lambda-\lambda_0\right)\kappa - \left(\lambda-\lambda_0\right)V_t''\left(k\right) - \kappa\lambda_0}{\left[\kappa\left(\varepsilon+1\right) - V_t''\left(k\right)\right]^2} \left(\frac{V_t''\left(k\right)}{2}\right)^2 \int \left(k'-k\right)^2 dF_t\left(k'\right),
\end{aligned}$$

where the second equality is because $V_t''\left(k\right)$ is a constant in k. Thus the sufficient and necessary condition for $\varepsilon_t\left(k\right)=1$ $\forall k$ to be an equilibrium is that $V_t''\left(k\right) \leq \frac{\lambda-2\lambda_0}{\lambda-\lambda_0}\kappa$.

[The largest equilibrium is either $\varepsilon_t\left(k\right)=0$ $\forall k$ or $\varepsilon_t\left(k\right)=1$ $\forall k$] Let $\varepsilon_t^{\max}\left(k\right)$ be the largest equilibrium search profile. Denote $\bar{\varepsilon}_t = \sup_k\left\{\varepsilon_t^{\max}\left(k\right)\right\}$ and $\underline{\varepsilon}_t = \inf_k\left\{\varepsilon_t^{\max}\left(k\right)\right\}$. We first prove $\bar{\varepsilon}_t = \underline{\varepsilon}_t$ by contradiction. Suppose $\bar{\varepsilon}_t > \underline{\varepsilon}_t$, then $\frac{\partial^2 S_t\left(k,k',\varepsilon,\varepsilon'\right) m\left(\varepsilon,\varepsilon'\right)}{\partial\varepsilon\partial\varepsilon'} > 0$ implies

$$\begin{aligned}
& \frac{\partial \int S_t\left(k,k',\varepsilon_t^{\max}\left(k\right),\bar{\varepsilon}_t\right) m\left(\varepsilon_t^{\max}\left(k\right),\bar{\varepsilon}_t\right) dF_t\left(k'\right)}{\partial \varepsilon} \\
\geq & \frac{\partial \int S_t\left(k,k',\varepsilon_t^{\max}\left(k\right),\varepsilon_t^{\max}\left(k'\right)\right) m\left(\varepsilon_t^{\max}\left(k\right),\varepsilon_t^{\max}\left(k'\right)\right) dF_t\left(k'\right)}{\partial \varepsilon} \geq 0
\end{aligned} \tag{C.7}$$

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 (C.7) implies that $0 \le (\lambda-\lambda_0)\,\kappa\,(\bar{\varepsilon}_t)^2 - (\lambda-\lambda_0)\,\bar{\varepsilon}_t V_t''(k) - \kappa\lambda_0 \le (\lambda-\lambda_0)\,\kappa - (\lambda-\lambda_0)\, V_t''(k) - \kappa\lambda_0$ for any k. Then we have

$$\frac{\partial \int S_t(k,k',\varepsilon,1)\, m(\varepsilon,1)\, dF_t(k')}{\partial \varepsilon} \ge 0 \text{ for any } k \text{ and } \varepsilon.$$

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,

$$\frac{\partial \int S_t(k,k',\varepsilon,1)\, m(\varepsilon,1)\, dF_t(k')}{\partial \varepsilon} > \frac{\partial \int S_t(k,k',\varepsilon,\hat{\varepsilon})\, m(\varepsilon,\hat{\varepsilon})\, dF_t(k')}{\partial \varepsilon} \propto \frac{\partial \int S_t(k,k',\hat{\varepsilon},\hat{\varepsilon})\, m(\varepsilon,\hat{\varepsilon})\, dF_t(k')}{\partial \varepsilon} = 0.$$

Thus there exists an equilibrium search profile where $\varepsilon_t(k) \equiv 1$, which is a contradiction. **Q.E.D.**

C.8 Proof of Proposition 5

Proof. Given $\{F_t\}$, the value function satisfying (8) is unique. Guess that

$$V_t(k) = -H_t k^2 + E_t k + D_t. \tag{C.8}$$

Then we have

$$V_t(k+q) - V_t(k) - \chi(\varepsilon,q) = \left[E_t - 2H_t k - (H_t + \kappa\varepsilon)\, q\right] q.$$

The bargaining solution thus solves

$$\begin{aligned}
q_t(k,k',\varepsilon,\varepsilon') &= \arg\max_q \left\{V_t(k+q) + V_t(k'-q) - \chi(\varepsilon,q) - \chi(\varepsilon',q)\right\}, \\
&= \arg\max_q \left\{-H_t(k+q)^2 - H_t(k'-q)^2 - \kappa(\varepsilon+\varepsilon')\, q^2\right\}, \\
&= \frac{H_t(k'-k)}{\kappa(\varepsilon+\varepsilon') + 2H_t},
\end{aligned}$$

and

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

Thus the bilateral Federal funds rate is

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

The trade surplus is given by

$$\begin{aligned}S_t\left(k,k',\varepsilon,\varepsilon'\right) &\equiv V_t\left[k+q_t\left(k,k',\varepsilon,\varepsilon'\right)\right]-V_t\left(k\right)-\chi\left[\varepsilon,q_t\left(k,k',\varepsilon,\varepsilon'\right)\right]\\ &\quad +V_t\left[k'-q_t\left(k,k',\varepsilon,\varepsilon'\right)\right]-V_t\left(k'\right)-\chi\left[\varepsilon',-q_t\left(k,k',\varepsilon,\varepsilon'\right)\right],\\ &= -H_t\left\{\left[k+\frac{H_t\left(k'-k\right)}{\kappa\left(\varepsilon+\varepsilon'\right)+2H_t}\right]^2+\left[k'-\frac{H_t\left(k'-k\right)}{\kappa\left(\varepsilon+\varepsilon'\right)+2H_t}\right]^2-k^2-k'^2\right\}\\ &\quad -\kappa\left(\varepsilon+\varepsilon'\right)\left[\frac{H_t\left(k'-k\right)}{\kappa\left(\varepsilon+\varepsilon'\right)+2H_t}\right]^2,\\ &= \frac{\left[H_t\left(k'-k\right)\right]^2}{\kappa\left(\varepsilon+\varepsilon'\right)+2H_t}.\end{aligned}$$

The equilibrium search profile is a fixed point function to the following functional:

$$\begin{aligned}\Gamma_t\left(\varepsilon_t\right)\left(k\right) &\equiv \arg\max_{\varepsilon\in[0,1]}\left\{\int S_t\left[k,k',\varepsilon,\varepsilon_t\left(k'\right)\right]m\left[\varepsilon,\varepsilon_t\left(k'\right)\right]dF_t\left(k'\right)\right\} \qquad \text{(C.9)}\\ &= \arg\max_{\varepsilon\in[0,1]}\left\{\int\frac{\left[H_t\left(k'-k\right)\right]^2}{\kappa\left[\varepsilon+\varepsilon_t\left(k'\right)\right]+2H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon\varepsilon_t\left(k'\right)+\lambda_0\right]dF_t\left(k'\right)\right\}\\ &= \arg\max_{\varepsilon\in[0,1]}\left\{\frac{\left(H_t\right)^2}{\kappa\left(\varepsilon+\varepsilon_t\right)+2H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon\varepsilon_t+\lambda_0\right]\int\left(k'-k\right)^2dF_t\left(k'\right)\right\}\end{aligned}$$

which only depends on H_t and F_t and, is independent of k and k'. The last equality is guaranteed by Proposition 4. Thus, we write $\Gamma_t\left(\varepsilon\right)\left(k\right)=\Gamma\left(\varepsilon;H_t\right)$, where the latter is given by (20). The equilibrium search intensity at t is the fixed point of $\Gamma\left(\varepsilon_t;H_t\right)$, which is any element of $\Omega\left(h\right)$. The HJB equation becomes

$$\begin{aligned}r\left[-H_tk^2+E_tk+D_t\right] &= -\dot{H}_tk^2+\dot{E}_tk+\dot{D}_t-a_2k^2+a_1k\\ &\quad +\frac{1}{4}\frac{H_t^2}{\kappa\varepsilon_t+H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2+\lambda_0\right]\int\left(k'-k\right)^2dF_t\left(k'\right).\end{aligned}$$

Matching the coefficients, we have

$$\begin{aligned}rH_t &= \dot{H}_t+a_2-\frac{1}{4}\frac{H_t^2}{\kappa\varepsilon_t+H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2+\lambda_0\right],\\ rE_t &= \dot{E}_t+a_1-\frac{1}{2}\frac{H_t^2K}{\kappa\varepsilon_t+H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2+\lambda_0\right],\\ rD_t &= \dot{D}_t+\frac{1}{4}\frac{H_t^2}{\kappa\varepsilon_t+H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2+\lambda_0\right]\int k'^2dF_t\left(k'\right),\end{aligned}$$

where the fact that $V_T\left(k\right)=-A_2k^2+A_1k$ implies the terminal conditions

$$H_T=A_2,\ E_T=A_1,\ D_T=0.$$

Q.E.D.

C.9 Proof of Proposition 6

Proof. Notice that the first-order condition of bank k's search intensity is

$$\begin{aligned}
& \frac{\partial \int S_t\left(k, k', \varepsilon, \varepsilon_t\right) m\left(\varepsilon, \varepsilon_t\right) dF_t\left(k'\right)}{\partial \varepsilon} \\
= & \frac{\left(\lambda-\lambda_0\right) \kappa\left(\varepsilon_t\right)^2+2\left(\lambda-\lambda_0\right) \varepsilon_t H_t-\kappa \lambda_0}{\left[\kappa\left(\varepsilon+\bar{\varepsilon}_t\right)+2 H_t\right]^2}\left(H_t\right)^2 \int\left(k'-k\right)^2 dF_t\left(k'\right).
\end{aligned}$$

Thus $\varepsilon=1$ if $\left(\lambda-\lambda_0\right) \kappa\left(\varepsilon_t\right)^2+2\left(\lambda-\lambda_0\right) \varepsilon_t H_t-\kappa \lambda_0>0$. It implies that the largest equilibrium satisfies

$$\Gamma_t\left(\boldsymbol{\varepsilon}_t\right)(k)\begin{cases}=1, \text{ if } H_t>\frac{\left[\lambda_0-\left(\lambda-\lambda_0\right)\right] \kappa}{2\left(\lambda-\lambda_0\right)}; \\ 0, \text{ otherwise.}\end{cases}$$

By rearranging the terms, we obtain the proposition. **Q.E.D.**

C.10 Proof of Lemma 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

$$\begin{aligned}
\dot{H}_t & = r H_t-a_2+\frac{\lambda}{4} \frac{H_t^2}{\kappa+H_t}=\frac{(4 r+\lambda) H_t^2+4\left(\kappa r-a_2\right) H_t-4 \kappa a_2}{4\left(\kappa+H_t\right)} \\
& = \frac{4 r+\lambda}{4\left(\kappa+H_t\right)}\left(H_t-\mu_1\right)\left(H_t-\mu_2\right),
\end{aligned} \tag{C.10}$$

where μ_1 and μ_2 are the zero point of formula $(4 r+\lambda) H_t^2+4\left(\kappa r-a_2\right) H_t-4 \kappa a_2=0$, and they are given by

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

Equation (C.10) can be written as

$$\begin{aligned}
\frac{4 r+\lambda}{4} dt & = \frac{\kappa+H_t}{\left(H_t-\mu_1\right)\left(H_t-\mu_2\right)} dH_t \\
& = \frac{\kappa+\mu_1}{\mu_1-\mu_2} \cdot \frac{1}{H_t-\mu_1} dH_t-\frac{\kappa+\mu_2}{\mu_1-\mu_2} \cdot \frac{1}{H_t-\mu_2} dH_t
\end{aligned}$$

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:

$$t = \tau_1 \left(H_t; A, u\right) = u - \frac{(\kappa + \mu_1) \log \left(\frac{A - \mu_1}{H_t - \mu_1}\right) - (\kappa + \mu_2) \log \left(\frac{A - \mu_2}{H_t - \mu_2}\right)}{\left(r + \frac{\lambda}{4}\right) (\mu_1 - \mu_2)}.$$

[Solve the ODE of H_t under $\varepsilon_t = 0$] If $\varepsilon_t = 0$, the law of motion of H_t is

$$\dot{H}_t = r H_t - a_2 + \frac{\lambda_0}{4} H_t = \left(r + \frac{\lambda_0}{4}\right) H_t - a_2. \tag{C.11}$$

Given a terminal value condition $H_u = A$, the solution to H_t is

$$H_t = J\left(t; A, u\right) = \frac{a_2}{r + \frac{\lambda_0}{4}} + \left(A - \frac{a_2}{r + \frac{\lambda_0}{4}}\right) e^{-\left(r + \frac{\lambda_0}{4}\right)(u - t)}.$$

Note that $J\left(t; A, u\right)$ is monotone in t. Thus we can get the inverse function:

$$t = \tau_2 \left(H; A, u\right) = u - \frac{1}{r + \frac{\lambda_0}{4}} \log \left(\frac{A - \frac{a_2}{r + \frac{\lambda_0}{4}}}{H - \frac{a_2}{r + \frac{\lambda_0}{4}}}\right) = u + \frac{1}{r + \frac{\lambda_0}{4}} \log \left(1 - \frac{H - A}{\frac{a_2}{r + \frac{\lambda_0}{4}} - A}\right).$$

[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** $\tau_1\left(H_t; A, u\right) > 0$**.** In this case, as t decreases from T, the equilibrium solution of ε_t and H_t is given by $\varepsilon_t = 1$ and $t = \tau_1\left(H_t; A_2, T\right)$. Moreover, according to equation (C.10), $\dot{H}_t\Big|_{H_t = \eta, \varepsilon_t = 1} > 0$ guarantees that H_t is decreasing as t goes from T to 0 before hitting η. According to the definition of $\tau_1\left(H_t; A, u\right)$, the time of H_t hitting η is $\tau_1\left(\eta; A_2, T\right)$. A positive $\tau_1\left(\eta; A_2, T\right)$ 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 $t < \tau_1\left(\eta; A_2, T\right)$. 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 $\tau_1\left(\eta; A_2, T\right) > 0$, the paths of ε_t and H_t are

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

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

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 (C.11), $\dot{H}_t\Big|_{H_t=\eta,\varepsilon_t=0} < 0$ 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.**

C.11 Proof of Lemma 3

Proof. Plug the closed-form solution (24) and $\varepsilon_t(k) \equiv \varepsilon_t$ into the KFE (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:

$$\dot{f}_t(k) = m(\varepsilon_t, \varepsilon_t)\left[2\left(1+\frac{\kappa\varepsilon_t}{H_t}\right)\int f_t\left[2\left(1+\frac{\kappa\varepsilon_t}{H_t}\right)k - \left(1+\frac{2\kappa\varepsilon_t}{H_t}\right)k'\right] f_t(k')\,dk' - f_t(k)\right]. \tag{C.12}$$

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 (C.12) is

$$\dot{f}_t^*(\nu) = m(\varepsilon_t, \varepsilon_t)\left[f_t^*\left(\frac{H_t}{2(H_t+\kappa\varepsilon_t)}\nu\right) f_t^*\left(\frac{H_t+2\kappa\varepsilon_t}{2(H_t+\kappa\varepsilon_t)}\nu\right) - f_t^*(\nu)\right]. \tag{C.13}$$

The PDE (C.13) 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 (C.13), we can obtain

$$\dot{f}_t^{*(n)}(\nu) = m(\varepsilon_t, \varepsilon_t)\left[\sum_{i=0}^{n} C_n^i \frac{(H_t)^{n-i}(H_t+2\kappa\varepsilon_t)^i}{2^n (H_t+\kappa\varepsilon_t)^n} f_t^{*(n-i)}\left(\frac{H_t}{2(H_t+\kappa\varepsilon_t)}\nu\right) f_t^{*(i)}\left(\frac{H_t+2\kappa\varepsilon_t}{2(H_t+\kappa\varepsilon_t)}\nu\right) - f_t^{*(n)}(\nu)\right]. \tag{C.14}$$

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 (33). **Q.E.D.**

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 (34).

Return reversal. The average Federal funds rate is

$$1+\varrho_t = e^{r(T+\Delta-t)}\left[E_t - 2H_t K\right], \tag{C.15}$$

then the difference between individual Federal funds rate and the average Federal funds rate is

$$\rho_t\left(k,k'\right) - \varrho_t = e^{r(T+\Delta-t)}\left(2K-k-k'\right)H_t.$$

Differentiating the rates with respect to time:

$$\dot{\varrho}_t = e^{r(T+\Delta-t)}\left(\dot{E}_t - 2K\dot{H}_t\right) - r\left(1+\varrho_t\right) = e^{r(T+\Delta-t)}\left(2a_2K - a_1\right), \tag{C.16}$$

$$\begin{aligned}
\dot{\rho}_t\left(k,k'\right) &= e^{r(T+\Delta-t)}\left[\dot{E}_t - \dot{H}_t\left(k+k'\right)\right] - r\left(1+\rho_t\left(k,k'\right)\right) \\
&= e^{r(T+\Delta-t)}\left[-a_1 + \left(k+k'\right)a_2 + \frac{2K-k-k'}{4}\frac{H_t^2}{\kappa\varepsilon_t + H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2 + \lambda_0\right]\right].
\end{aligned}$$

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

$$\begin{aligned}
&\Delta_{\rho,t}\left(k,q\right) \\
\equiv\ &\int\rho_t\left(k+q,k'\right)dF_t\left(k'\right) - \rho_t\left(k,q\right) \\
=\ &\int e^{r(T+\Delta-t)}\left[E_t - H_t\left(k+q+k'\right)\right]dF_t\left(k'\right) - e^{r(T+\Delta-t)}\left[E_t - H_t\left(k+k+\frac{2\left(\kappa\varepsilon_t + H_t\right)}{H_t}q\right)\right] \\
=\ &e^{r(T+\Delta-t)}\left[-H_t\left(k+q+K\right) + 2kH_t + 2\left(\kappa\varepsilon_t + H_t\right)q\right] \\
=\ &e^{r(T+\Delta-t)}\left[-H_t\left(K-k\right) + \left(2\kappa\varepsilon_t + H_t\right)q\right].
\end{aligned}$$

Thus the intermediation markup is given by taking $\Delta_{\rho,t}\left(k,q\right)$ differentiation with respect to q.

Unilization rate of trade opportunities. By definition,

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

Extensive margins. We provide a heuristic approach to derive the dynamics of the extensive margins. Let Δ be a small time length, and denote $m_t \equiv m\left(\varepsilon_t, \varepsilon_t\right)$ as the equilibrium matching rate. Then by definition,

$$
1 - P_t^{\emptyset}\left(k\right) = \left(1 - \Delta \cdot m_t\right)\left[1 - P_{t+\Delta}^{\emptyset}\left(k\right)\right] + \Delta \cdot m_t \cdot 0,
$$

where $1 - P_t^{\emptyset}\left(k\right)$ 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

$$
\dot{P}_t^{\emptyset}\left(k\right) = \lim_{\Delta\to 0} \frac{P_{t+\Delta}^{\emptyset}\left(k\right) - P_t^{\emptyset}\left(k\right)}{\Delta} = -m_t\left[1 - P_t^{\emptyset}\left(k\right)\right].
$$

The evolution of $P_t^b\left(k\right)$ and $P_t^s\left(k\right)$ can be derived similarly as follows.

$$
1 - P_t^b\left(k\right) = \left(1 - \Delta \cdot m_t\right)\left[1 - P_{t+\Delta}^b\left(k\right)\right] + \Delta \cdot m_t \int_{k'\leq k} \left[1 - P_{t+\Delta}^b\left(k + q_t\left(k, k'\right)\right)\right] dF_t\left(k'\right),
$$

$$
1 - P_t^s\left(k\right) = \left(1 - \Delta \cdot m_t\right)\left[1 - P_{t+\Delta}^s\left(k\right)\right] + \Delta \cdot m_t \int_{k'\geq k} \left[1 - P_{t+\Delta}^s\left(k + q_t\left(k, k'\right)\right)\right] dF_t\left(k'\right).
$$

Take $\Delta \to 0$ gives

$$
\dot{P}_t^b\left(k\right) = -m_t\left[1 - F_t\left(k\right)\right]\left[1 - P_t^b\left(k\right)\right] - m_t \int_{k'\leq k} \left[P_t^b\left(k + q_t\left(k, k'\right)\right) - P_t^b\left(k\right)\right] dF_t\left(k'\right),
$$

$$
\dot{P}_t^s\left(k\right) = -m_t F_t\left(k\right)\left[1 - P_t^s\left(k\right)\right] - m_t \int_{k'\geq k} \left[P_t^s\left(k + q_t\left(k, k'\right)\right) - P_t^s\left(k\right)\right] dF_t\left(k'\right).
$$

Then the evolution of $P_t^{int}\left(k\right)$ is

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

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,

$$\begin{aligned}
Q_t(k) &= \Delta \cdot m(\varepsilon_t, \varepsilon_t) \cdot \left[\int_{k'} \left|q_t(k, k')\right| dF_t(k') + \int_{k'} Q_{t+\Delta}\left(k + q_t(k, k')\right) dF_t(k')\right] \\
&\quad + \left[1 - \Delta \cdot m(\varepsilon_t, \varepsilon_t)\right] Q_{t+\Delta}(k).
\end{aligned}$$

Thus the aggregate absolute trades is given by

$$\begin{aligned}
Q_t &= \int Q_t(k)\, dF_t(k) \\
&= \Delta \cdot m(\varepsilon_t, \varepsilon_t) \cdot \int_k \int_{k'} \left|q_t(k, k')\right| dF_t(k')\, dF_t(k) \\
&\quad + \int_k \left\{\Delta \cdot m(\varepsilon_t, \varepsilon_t) \cdot \int_{k'} Q_{t+\Delta}\left(k + q_t(k, k')\right) dF_t(k') + \left[1 - \Delta \cdot m(\varepsilon_t, \varepsilon_t)\right] Q_{t+\Delta}(k)\right\} dF_t(k) \\
&= \Delta \cdot m(\varepsilon_t, \varepsilon_t) \cdot \int_k \int_{k'} \left|q_t(k, k')\right| dF_t(k')\, dF_t(k) + Q_{t+\Delta},
\end{aligned}$$

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:

$$\begin{aligned}
\dot{Q}_t(k) &= \lim_{\Delta \to 0} \frac{Q_{t+\Delta}(k) - Q_t(k)}{\Delta} \\
&= -m(\varepsilon_t, \varepsilon_t) \cdot \left[\int_{k'} \left|q_t(k, k')\right| dF_t(k') + \int_{k'} Q_{t+\Delta}\left(k + q_t(k, k')\right) dF_t(k')\right] + m(\varepsilon_t, \varepsilon_t)\, Q_t(k),
\end{aligned}$$

and

$$\begin{aligned}
\dot{Q}_t &= \lim_{\Delta \to 0} \frac{Q_{t+\Delta} - Q_t}{\Delta} \\
&= -m(\varepsilon_t, \varepsilon_t) \cdot \int_k \int_{k'} \left|q_t(k, k')\right| dF_t(k')\, dF_t(k) \\
&= -m(\varepsilon_t, \varepsilon_t) \frac{H_t}{2(\kappa\varepsilon_t + H_t)} \int_k \int_{k'} \left|k' - k\right| dF_t(k')\, dF_t(k).
\end{aligned}$$

This implies

$$Q = \int_0^T m(\varepsilon_t, \varepsilon_t) \frac{H_t}{2(\kappa\varepsilon_t + H_t)} \left(\int_k \int_{k'} \left|k' - k\right| dF_t(k')\, dF_t(k)\right) dt. \tag{C.17}$$

Second, for the individual net Federal funds purchase, it satisfies

$$\begin{aligned}
L_t(k) &= \Delta \cdot m_t \int \frac{H_t(k' - k)}{2(\kappa\varepsilon_t + H_t)} dF_t(k') + \Delta \cdot m_t \int L_{t+\Delta}\left(k + q_t(k, k')\right) dF_t(k') \\
&\quad + (1 - \Delta \cdot m_t)\, L_{t+\Delta}(k) \\
&= \Delta \cdot m_t \frac{H_t(K - k)}{2(\kappa\varepsilon_t + H_t)} + \Delta \cdot m_t \int L_{t+\Delta}\left(k + q_t(k, k')\right) dF_t(k') \\
&\quad + (1 - \Delta \cdot m_t)\, L_{t+\Delta}(k).
\end{aligned}$$

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

$$L = \int |L_0(k)|\, dF_0(k) = \left\{1 - \exp\left[-\int_0^T \frac{m_z H_z}{2(\kappa\varepsilon_z + H_z)}dz\right]\right\}\int |K - k|\, dF_0(k). \tag{C.18}$$

Federal funds rate. The average Federal funds rate at t is given by equation (C.15). It satisfies the ODE (C.16) with terminal condition $1 + \varrho_T = e^{r\Delta}[A_1 - 2A_2K]$, which has the following closed-form solution:

$$\begin{aligned}1 + \varrho_t &= e^{r\Delta}(A_1 - 2A_2K) - \frac{2a_2K - a_1}{r}\left[e^{r(T+\Delta-t)} - e^{r\Delta}\right]\\ &= e^{r\Delta}\left[1 + \gamma + \frac{(k_+ - 1)i^{DW} - (k_- - 1)i^{ER}}{k_+ - k_-}\right] - \frac{2a_2K - a_1}{r}\left[e^{r(T+\Delta-t)} - e^{r\Delta}\right]\\ &= e^{r\Delta}\left[1 + \gamma + i^{ER} + \frac{k_+ - 1}{k_+ - k_-}\Delta i\right] - \frac{2a_2K - a_1}{r}\left[e^{r(T+\Delta-t)} - e^{r\Delta}\right].\end{aligned}$$

C.13 Proof of Proposition 7

Proof.

Comparative statics for the length of search. Note that the length of search in this case is given by

$$\tau_2(\eta; A_2, T) = T + \frac{1}{r + \frac{\lambda_0}{4}}\log\left(1 - \frac{\eta - A_2}{\frac{a_2}{r + \frac{\lambda_0}{4}} - A_2}\right). \tag{C.19}$$

Then the first column of table in the proposition is given by differentiating (C.19). That is,

$$\frac{\partial\tau_2(\eta; A_2, T)}{\partial i^{ER}} = \frac{1}{r + \frac{\lambda_0}{4}}\frac{1}{1 - \frac{\eta - A_2}{\frac{a_2}{r + \frac{\lambda_0}{4}} - A_2}}\frac{\frac{a_2}{r + \frac{\lambda_0}{4}} - \eta}{\left(\frac{a_2}{r + \frac{\lambda_0}{4}} - A_2\right)^2}\left[-\frac{1}{2K(k_+ - k_-)}\right] < 0,$$

$$\frac{\partial\tau_2(\eta;A_2,T)}{\partial i^{DW}} = \frac{1}{r+\frac{\lambda_0}{4}}\frac{1}{1-\frac{\eta-A_2}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}}\frac{\frac{a_2}{r+\frac{\lambda_0}{4}}-\eta}{\left(\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2\right)^2}\left[\frac{1}{2K(k_+-k_-)}\right] > 0,$$

$$\frac{\partial\tau_2(\eta;A_2,T)}{\partial K} = \frac{1}{r+\frac{\lambda_0}{4}}\frac{1}{1-\frac{\eta-A_2}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}}\frac{\frac{a_2}{r+\frac{\lambda_0}{4}}-\eta}{\left(\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2\right)^2}\left[-\frac{i^{DW}-i^{ER}}{2K^2(k_+-k_-)}\right] < 0,$$

$$\frac{\partial\tau_2(\eta;A_2,T)}{\partial\kappa} = \frac{1}{r+\frac{\lambda_0}{4}}\frac{1}{1-\frac{\eta-A_2}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}}\left\{-\frac{1}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}\left[\frac{\lambda}{2(\lambda-\lambda_0)}-1\right]\right\} < 0,$$

$$\begin{aligned}\frac{\partial\tau_2(\eta;A_2,T)}{\partial\lambda_0} &= -\frac{1}{4\left(r+\frac{\lambda_0}{4}\right)^2}\log\left(1-\frac{\eta-A_2}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}\right)\\ &\quad+\frac{1}{r+\frac{\lambda_0}{4}}\frac{1}{1-\frac{\eta-A_2}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}}\left\{-\frac{\frac{\kappa\lambda}{2(\lambda-\lambda_0)^2}\left(\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2\right)+(\eta-A_2)\frac{a_2}{4\left(r+\frac{\lambda_0}{4}\right)^2}}{\left(\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2\right)^2}\right\}\\ &< 0,\end{aligned}$$

$$\frac{\partial\tau_2(\eta;A_2,T)}{\partial\lambda} = \frac{1}{r+\frac{\lambda_0}{4}}\frac{1}{1-\frac{\eta-A_2}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}}\frac{1}{\frac{a_2}{r+\frac{\lambda_0}{4}}-A_2}\frac{\kappa\lambda_0}{2(\lambda-\lambda_0)^2} > 0.$$

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 (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

$$\begin{aligned}
\frac{\partial \Phi}{\partial A_2} &= \left(\frac{\lambda}{2} \frac{H_{\tau_2}}{\kappa + H_{\tau_2}} - \frac{\lambda_0}{2}\right) \frac{\partial \tau_2}{\partial A_2} + \frac{\lambda}{2} \int_0^{\tau_2} \frac{\kappa}{(\kappa + H_t)^2} \frac{\partial H_t}{\partial A_2} dt \\
&= -\frac{\lambda - \lambda_0}{2} \frac{\partial \tau_2}{\partial A_2} + \frac{\lambda}{2} \int_0^{\tau_2} \frac{\kappa}{(\kappa + H_t)^2} \left(-\frac{\partial \tau_2 / \partial A_2}{\partial \tau_1(H_t; \eta, \tau_2) / \partial H_t}\right) dt \\
&= -\frac{\lambda - \lambda_0}{2} \frac{\partial \tau_2}{\partial A_2} + \frac{\lambda}{2} \frac{\partial \tau_2}{\partial A_2} \int_0^{\tau_2} \frac{\kappa}{(\kappa + H_t)^2} \left(-\dot{H}_t\right) dt \\
&= \left\{-\frac{\lambda - \lambda_0}{2} + \frac{\lambda}{2} \left[\frac{\kappa}{(\kappa + H_{\tau_2})} - \frac{\kappa}{(\kappa + H_0)}\right]\right\} \frac{\partial \tau_2}{\partial A_2} \\
&= \left(\frac{\lambda}{2} \frac{H_0}{H_0 + \kappa} - \frac{\lambda_0}{2}\right) \frac{\partial \tau_2}{\partial A_2}.
\end{aligned}$$

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 (21) and (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

$$\begin{aligned} \rho_t(k,k') &= e^{r(T+\Delta)}\left\{e^{-rT}\left[E_T - (k+k')H_T\right]\right. \\ &\quad \left. + \int_t^T e^{-rs}\left[a_1 - (k+k')a_2 + \frac{k+k'-2K}{4}\cdot\frac{H_s^2\left[(\lambda-\lambda_0)\varepsilon_s^2+\lambda_0\right]}{\kappa\varepsilon_s + H_s}\right]ds\right\}, \end{aligned} \tag{C.20}$$

where

$$E_T - (k+k')H_T = 1 + \frac{k_+ i^{DW} - k_- i^{ER}}{k_+ - k_-} + \gamma - \frac{i^{DW} - i^{ER}}{2K(k_+ - k_-)}(k+k'). \tag{C.21}$$

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 $t = \tau_1 (H_t; \eta, \tau_2 (\eta; A_2, T))$. The implicit function theorem implies that

$$\begin{aligned}\frac{\partial}{\partial i^{DW}} \left(\frac{\lambda H_t^2}{\kappa + H_t} \right) &= \lambda \frac{H_t (2\kappa + H_t)}{(\kappa + H_t)^2} \frac{\partial H_t}{\partial i^{DW}} \\ &= \lambda \frac{H_t (2\kappa + H_t)}{(\kappa + H_t)^2} \left(-\dot{H}_t \right) \frac{\partial \tau_2 (\eta; A_2, T)}{\partial A_2} \frac{\partial A_2}{\partial i^{DW}}.\end{aligned}$$

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 $M(u)$:

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

which is independent of time t. Thus we can rewrite $\frac{\partial \rho_t (k, k')}{\partial i^{DW}}$ as

$$\frac{\partial \rho_t (k, k')}{\partial i^{DW}} \propto e^{r(T+\Delta)} \left\{ e^{-rT} \frac{k_+ - (k + k')/2K}{k_+ - k_-} + \frac{(k + k')/2K - 1}{k_+ - k_-} \left[M(\tau_2 - t) + e^{-rT} \left(1 - e^{-\frac{\lambda_0}{4}(T - \tau_2)} \right) \right] \right\}.$$

It follows that $\frac{\partial \rho_t (k, k')}{\partial i^{DW}} < 0$ iff

$$k + k' > 2K \frac{k_+ - 1 + e^{-\frac{\lambda_0}{4}(T - \tau_2)} - M(\tau_2 - t) e^{rT}}{e^{-\frac{\lambda_0}{4}(T - \tau_2)} - M(\tau_2 - t) e^{rT}}.$$

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

$$\frac{\partial \rho_t (k, k')}{\partial \lambda_0} = \frac{k + k' - 2K}{4} e^{r(T+\Delta)} \int_t^T e^{-rs} \cdot \frac{\partial [\lambda_0 H_s]}{\partial \lambda_0} ds.$$

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

$$\begin{aligned}
\frac{\partial \rho_t (k, k')}{\partial \lambda_0} &= e^{r(T+\Delta)} \Bigg\{ -e^{-r\tau_2} \left[a_1 - (k + k') \, a_2 + \frac{k + k' - 2K}{4} \cdot \lambda_0 \eta \right] \frac{\partial \tau_2}{\partial \lambda_0} \\
&\quad + e^{-r\tau_2} \left[a_1 - (k + k') \, a_2 + \frac{k + k' - 2K}{4} \cdot \frac{\lambda \eta^2}{\kappa + \eta} \right] \frac{\partial \tau_2}{\partial \lambda_0} \\
&\quad + \int_{\tau_2}^T e^{-rs} \frac{k + k' - 2K}{4} \cdot \frac{\partial [\lambda_0 H_s]}{\partial \lambda_0} ds \\
&\quad + \int_t^{\tau_2} e^{-rs} \frac{k + k' - 2K}{4} \cdot \frac{\partial}{\partial \lambda_0} \left[\frac{\lambda H_s^2}{\kappa + H_s} \right] ds \Bigg\} \\
&= \frac{k + k' - 2K}{4} e^{r(T+\Delta-\tau_2)} \times \\
&\quad \left\{ \int_{\tau_2}^T e^{-r(s-\tau_2)} \frac{\partial [\lambda_0 H_s]}{\partial \lambda_0} ds + \int_t^{\tau_2} e^{r(\tau_2 - s)} \frac{\partial}{\partial \lambda_0} \left[\frac{\lambda H_s^2}{\kappa + H_s} \right] ds - (\lambda - \lambda_0) \, \eta \frac{\partial \tau_2}{\partial \lambda_0} \right\}.
\end{aligned}$$

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.**

C.14 Proof of Proposition 8

Proof.

Comparative statics of the length of search. The length of search in this case is given by

$$T - \tau_1 (\eta; A_2, T) = \frac{(\kappa + \mu_1) \log \left(\frac{A_2 - \mu_1}{\eta - \mu_1} \right) - (\kappa + \mu_2) \log \left(\frac{A_2 - \mu_2}{\eta - \mu_2} \right)}{\left(r + \frac{\lambda}{4} \right) (\mu_1 - \mu_2)}. \tag{C.22}$$

Then the first column of table in the proposition is given by differentiating (C.22). Note that in this case, we have $\mu_1 < 0 < \mu_2 < \eta \leq A_2$. Therefore, we obtain

$$\frac{\partial [T - \tau_1 (\eta; A_2, T)]}{\partial i^{ER}} = \frac{(\kappa + A_2)}{\left(r + \frac{\lambda}{4} \right) (A_2 - \mu_1) (A_2 - \mu_2)} \left[-\frac{1}{2K (k_+ - k_-)} \right] < 0,$$

$$\frac{\partial [T - \tau_1 (\eta; A_2, T)]}{\partial i^{DW}} = \frac{(\kappa + A_2)}{\left(r + \frac{\lambda}{4} \right) (A_2 - \mu_1) (A_2 - \mu_2)} \left[\frac{1}{2K (k_+ - k_-)} \right] > 0,$$

$$\frac{\partial [T - \tau_1 (\eta; A_2, T)]}{\partial K} = \frac{(\kappa + A_2)}{\left(r + \frac{\lambda}{4} \right) (A_2 - \mu_1) (A_2 - \mu_2)} \left[-\frac{i^{DW} - i^{ER}}{2K^2 (k_+ - k_-)} \right] < 0,$$

$$\frac{\partial [T - \tau_1 (\eta; A_2, T)]}{\partial \lambda_0} = -\frac{\kappa + \eta}{(\eta - \mu_1) (\eta - \mu_2) \left(r + \frac{\lambda}{4} \right)} \frac{\kappa \lambda}{2 (\lambda - \lambda_0)^2} < 0,$$

For the comparative statics w.r.t. κ, we define $\tilde{q}_t \equiv \frac{H_t}{\kappa+H_t}$. Following the derivations in 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 $t > \tau_1(\eta; A_2, T)$, 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 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 $t \in [\tau_1(\eta; A_2, T), T]$. 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 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 $t \in \hat{h}_t \in [\tau_1(\eta; A_2, T), T]$, $\hat{h}_t \in \left[\lambda_0 - \lambda, \lambda\frac{A_2}{\kappa + A_2} - \lambda_0\right] \subset (-\lambda_0, \lambda - \lambda_0)$, $\dot{\hat{h}}_t > 0$, $\frac{\partial \dot{\hat{h}}_t}{\partial \lambda}\Big|_{\hat{h}_t = -\lambda_0} = -\frac{a_2}{\kappa} < 0$, $\frac{\partial \dot{\hat{h}}_t}{\partial \lambda}\Big|_{\hat{h}_t = \lambda - \lambda_0} = r + \frac{\lambda}{4} > 0$, 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 C.13. For the comparative statics of $\rho_t(k,k')$ w.r.t. i^{DW}, we can take derivative to C.20. When $t > \tau_1(\eta; A_2, T)$, we have

$$\begin{aligned}
\frac{\partial \rho_t(k,k')}{\partial i^{DW}} &= e^{r(T+\Delta)}\left\{e^{-rT}\left[\frac{k_+}{k_+ - k_-} - \frac{k+k'}{2K(k_+ - k_-)}\right]\right. \\
&\quad \left. + \frac{\lambda}{4}\left(k+k'-2K\right)\int_t^T e^{-rs}\frac{\partial}{\partial i^{DW}}\left[\frac{H_s^2}{\kappa + H_s}\right]ds\right\} \\
&= e^{r(T+\Delta)}\left\{e^{-rT}\left[\frac{k_+}{k_+ - k_-} - \frac{k+k'}{2K(k_+ - k_-)}\right]\right. \\
&\quad \left. + \left(k+k'-2K\right)\frac{\partial A_2}{\partial i^{DW}}\frac{\lambda}{4}\frac{\partial \tau_1(\eta; A_2, T)}{\partial A_2}\int_t^T e^{-rs}\frac{H_s(2\kappa + H_s)}{(\kappa + H_s)^2}(-dH_s)\right\} \\
&= e^{r(T+\Delta)}\left\{e^{-rT}\left[\frac{k_+}{k_+ - k_-} - \frac{k+k'}{2K(k_+ - k_-)}\right]\right. \\
&\quad \left. + \frac{(k+k'-2K)}{2K(k_+ - k_-)}\tilde{M}(T-t)\right\},
\end{aligned}$$

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

$$\begin{aligned}
\rho_t\left(k,k'\right) &= e^{r(T+\Delta)}\left\{e^{-rT}\left[E_T - \left(k+k'\right)H_T\right]\right. \\
&\quad + \int_{\tau_1}^T e^{-rs}\left[a_1 - \left(k+k'\right)a_2 + \frac{k+k'-2K}{4}\cdot\frac{\lambda H_s^2}{\kappa + H_s}\right]ds \\
&\quad \left. + \int_t^{\tau_1} e^{-rs}\left[a_1 - \left(k+k'\right)a_2 + \frac{k+k'-2K}{4}\cdot\lambda_0 H_s\right]ds\right\},
\end{aligned}$$

and

$$
\begin{aligned}
\frac{\partial \rho_t\left(k,k'\right)}{\partial i^{DW}} &= e^{r(T+\Delta)}\left\{e^{-rT}\left[\frac{k_+}{k_+-k_-}-\frac{k+k'}{2K\left(k_+-k_-\right)}\right]+\frac{\left(k+k'-2K\right)}{2K\left(k_+-k_-\right)}\tilde{M}\left(T-\tau_1\right)\right. \\
&\quad -e^{-r\tau_1}\left[a_1-\left(k+k'\right)a_2+\frac{k+k'-2K}{4}\cdot\frac{\lambda\eta^2}{\kappa+\eta}\right]\frac{\partial \tau_1}{\partial A_2}\frac{\partial A_2}{i^{DW}} \\
&\quad \left.+e^{-r\tau_1}\left[a_1-\left(k+k'\right)a_2+\frac{k+k'-2K}{4}\cdot\lambda_0\eta\right]\frac{\partial \tau_1}{\partial A_2}\frac{\partial A_2}{i^{DW}}\right\} \\
&= e^{r(T+\Delta)}\left\{e^{-rT}\left[\frac{k_+}{k_+-k_-}-\frac{k+k'}{2K\left(k_+-k_-\right)}\right]+\frac{\left(k+k'-2K\right)}{2K\left(k_+-k_-\right)}\tilde{M}\left(T-\tau_1\right)\right. \\
&\quad \left.+\frac{k+k'-2K}{2K\left(k_+-k_-\right)}e^{-r\tau_1}\frac{\left(\lambda-\lambda_0\right)}{4}\eta\frac{\partial \tau_1}{\partial A_2}\right\}.
\end{aligned}
$$

Thus we have $\frac{\partial \rho_t(k,k')}{\partial i^{DW}} < 0$ iff

$$
k+k' > 2K\frac{k_+-\tilde{M}\left(T-\tau_1\right)e^{rT}-e^{r(T-\tau_1)}\frac{(\lambda-\lambda_0)\eta}{4}\frac{\partial \tau_1}{\partial A_2}}{1-\tilde{M}\left(T-\tau_1\right)e^{rT}-e^{r(T-\tau_1)}\frac{(\lambda-\lambda_0)\eta}{4}\frac{\partial \tau_1}{\partial A_2}}.
$$

We can also derive the comparative statics w.r.t. i^{ER} in a similar way. The result is that when $t > \tau_1\left(\eta; A_2, T\right)$, $\frac{\partial \rho_t(k,k')}{\partial i^{ER}} < 0$ iff

$$
k+k' < 2K\frac{k_- -\tilde{M}\left(T-t\right)e^{rT}}{1-\tilde{M}\left(T-t\right)e^{rT}}.
$$

When $t < \tau_1\left(\eta; A_2, T\right)$, $\frac{\partial \rho_t(k,k')}{\partial i^{ER}} < 0$ iff

$$
k+k' < 2K\frac{k_- -\tilde{M}\left(T-\tau_1\right)e^{rT}-e^{r(T-\tau_1)}\frac{(\lambda-\lambda_0)\eta}{4}\frac{\partial \tau_1}{\partial A_2}}{1-\tilde{M}\left(T-\tau_1\right)e^{rT}-e^{r(T-\tau_1)}\frac{(\lambda-\lambda_0)\eta}{4}\frac{\partial \tau_1}{\partial A_2}}.
$$

For the comparative statics w.r.t. K, note that when $t > \tau_1$, we have

$$
\begin{aligned}
\frac{\partial \rho_t\left(k,k'\right)}{\partial K} &= e^{r(T+\Delta)}\left\{e^{-rT}\frac{\left(i^{DW}-i^{ER}\right)\left(k+k'\right)}{2K^2\left(k_+-k_-\right)}-\frac{\lambda}{2}\int_t^T e^{-rs}\frac{H_s^2}{\kappa+H_s}ds\right. \\
&\quad \left.+\frac{\lambda}{4}\left(k+k'-2K\right)\int_t^T e^{-rs}\frac{\partial}{\partial K}\left[\frac{H_s^2}{\kappa+H_s}\right]ds\right\} \\
&= e^{r(T+\Delta)}\left\{e^{-rT}\frac{\left(i^{DW}-i^{ER}\right)\left(k+k'\right)}{2K^2\left(k_+-k_-\right)}-\frac{\lambda}{2}\int_t^T e^{-rs}\frac{H_s^2}{\kappa+H_s}ds\right. \\
&\quad \left.-\frac{\left(i^{DW}-i^{ER}\right)\left(k+k'-2K\right)}{2K^2\left(k_+-k_-\right)}\tilde{M}\left(T-t\right)\right\}.
\end{aligned}
$$

Thus $\frac{\partial \rho_t(k,k')}{\partial K} < 0$ iff

$$
k+k' < 2K\frac{\frac{\lambda}{4A_2}\int_t^T e^{r(T-s)}\frac{H_s^2}{\kappa+H_s}ds-\tilde{M}\left(T-t\right)e^{rT}}{1-\tilde{M}\left(T-t\right)e^{rT}}.
$$

When $t < \tau_1$, we have

$$\begin{aligned}
\frac{\partial \rho_t(k,k')}{\partial K} &= e^{r(T+\Delta)}\left\{e^{-rT}\frac{\left(i^{DW}-i^{ER}\right)(k+k')}{2K^2(k_+-k_-)} - \frac{1}{2}\int_t^T e^{-rs}\frac{\left[(\lambda-\lambda_0)\varepsilon_s^2+\lambda_0\right]H_s^2}{\kappa\varepsilon_s+H_s}ds\right.\\
&\quad -\frac{\left(i^{DW}-i^{ER}\right)(k+k'-2K)}{2K^2(k_+-k_-)}\tilde{M}(T-\tau_1)\\
&\quad -e^{-r\tau_1}\left[a_1-\left(k+k'\right)a_2+\frac{k+k'-2K}{4}\cdot\frac{\lambda\eta^2}{\kappa+\eta}\right]\frac{\partial\tau_1}{\partial A_2}\frac{\partial A_2}{K}\\
&\quad \left.+e^{-r\tau_1}\left[a_1-\left(k+k'\right)a_2+\frac{k+k'-2K}{4}\cdot\lambda_0\eta\right]\frac{\partial\tau_1}{\partial A_2}\frac{\partial A_2}{K}\right\}\\
&= e^{r(T+\Delta)}\left\{e^{-rT}\frac{\left(i^{DW}-i^{ER}\right)(k+k')}{2K^2(k_+-k_-)} - \frac{1}{2}\int_t^T e^{-rs}\frac{\left[(\lambda-\lambda_0)\varepsilon_s^2+\lambda_0\right]H_s^2}{\kappa\varepsilon_s+H_s}ds\right.\\
&\quad -\frac{\left(i^{DW}-i^{ER}\right)(k+k'-2K)}{2K^2(k_+-k_-)}\tilde{M}(T-\tau_1)\\
&\quad \left.-\frac{\left(i^{DW}-i^{ER}\right)(k+k'-2K)}{2K^2(k_+-k_-)}e^{-r\tau_1}\frac{(\lambda-\lambda_0)}{4}\eta\frac{\partial\tau_1}{\partial A_2}\right\}.
\end{aligned}$$

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 $t > \tau_1$,

$$\begin{aligned}
\rho_t\left(k,k'\right) &= e^{r(T+\Delta)}\left\{e^{-rT}\left[E_T-\left(k+k'\right)H_T\right]\right.\\
&\quad \left.+\int_t^T e^{-rs}\left[a_1-\left(k+k'\right)a_2+\frac{k+k'-2K}{4}\cdot\frac{\lambda H_s^2}{\kappa+H_s}\right]ds\right\},
\end{aligned}$$

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

$$\begin{aligned}
\rho_t\left(k,k'\right) &= e^{r(T+\Delta)}\left\{e^{-rT}\left[E_T-\left(k+k'\right)H_T\right]\right.\\
&\quad +\int_{\tau_1}^T e^{-rs}\left[a_1-\left(k+k'\right)a_2+\frac{k+k'-2K}{4}\cdot\frac{\lambda H_s^2}{\kappa+H_s}\right]ds\\
&\quad \left.+\int_t^{\tau_1} e^{-rs}\left[a_1-\left(k+k'\right)a_2+\frac{k+k'-2K}{4}\cdot\lambda_0 H_s\right]ds\right\},
\end{aligned}$$

and

$$
\begin{aligned}
\frac{\partial \rho_t(k,k')}{\partial \lambda} &= e^{r(T+\Delta)}\Bigg\{-e^{-r\tau_1}\left[a_1-\left(k+k'\right)a_2+\frac{k+k'-2K}{4}\cdot\frac{\lambda\eta^2}{\kappa+\eta}\right]\frac{\partial \tau_1}{\partial \lambda} \\
&\quad + e^{-r\tau_1}\left[a_1-\left(k+k'\right)a_2+\frac{k+k'-2K}{4}\cdot\lambda_0\eta\right]\frac{\partial \tau_1}{\partial \lambda} \\
&\quad + \int_{\tau_1}^{T} e^{-rs}\frac{k+k'-2K}{4}\cdot\frac{\partial}{\partial \lambda}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds \\
&\quad + \int_{t}^{\tau_1} e^{-rs}\frac{k+k'-2K}{4}\cdot\frac{\partial\left[\lambda_0 H_s\right]}{\partial \lambda}ds\Bigg\} \\
&= \frac{k+k'-2K}{4}e^{r(T+\Delta)}\Bigg\{\int_{\tau_1}^{T} e^{-r(s-\tau_1)}\frac{\partial}{\partial \lambda}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds+\int_t^{\tau_1} e^{r(\tau_1-s)}\frac{\partial\left[\lambda_0 H_s\right]}{\partial \lambda}ds \\
&\quad + \left(\lambda-\lambda_0\right)\eta\frac{\partial \tau_1}{\partial \lambda}\Bigg\},
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{\partial \rho_t(k,k')}{\partial \lambda} &= e^{r(T+\Delta)}\Bigg\{-e^{-r\tau_1}\left[a_1-\left(k+k'\right)a_2+\frac{k+k'-2K}{4}\cdot\frac{\lambda\eta^2}{\kappa+\eta}\right]\frac{\partial \tau_1}{\partial \kappa} \\
&\quad + e^{-r\tau_1}\left[a_1-\left(k+k'\right)a_2+\frac{k+k'-2K}{4}\cdot\lambda_0\eta\right]\frac{\partial \tau_1}{\partial \kappa} \\
&\quad + \int_{\tau_1}^{T} e^{-rs}\frac{k+k'-2K}{4}\cdot\frac{\partial}{\partial \kappa}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds \\
&\quad + \int_{t}^{\tau_1} e^{-rs}\frac{k+k'-2K}{4}\cdot\frac{\partial\left[\lambda_0 H_s\right]}{\partial \kappa}ds\Bigg\} \\
&= \frac{k+k'-2K}{4}e^{r(T+\Delta)}\Bigg\{\int_{\tau_1}^{T} e^{-r(s-\tau_1)}\frac{\partial}{\partial \kappa}\left[\frac{\lambda H_s^2}{\kappa+H_s}\right]ds+\int_t^{\tau_1} e^{r(\tau_1-s)}\frac{\partial\left[\lambda_0 H_s\right]}{\partial \kappa}ds \\
&\quad + \left(\lambda-\lambda_0\right)\eta\frac{\partial \tau_1}{\partial \kappa}\Bigg\},
\end{aligned}
$$

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 $\frac{\partial \rho_t(k,k')}{\partial \lambda}<0$ 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.**

C.15 Proof of Proposition 9

Proof. The proposition is a restatement of equations (12) and (14). **Q.E.D.**

C.16 Proof of Lemma 4

Proof. The proof follows Lemma 2. **Q.E.D.**

C.17 Proof of Proposition 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

$$
\begin{aligned}
\dot{H}_t &= rH_t - a_2 + \frac{1}{4}\frac{H_t^2\left[(\lambda-\lambda_0)\cdot 1\{H_t \geq \eta\}+\lambda_0\right]}{\kappa\cdot 1\{H_t\geq\eta\}+H_t}, \\
\dot{H}_t^p &= rH_t^p - a_2 + \frac{1}{2}\frac{(H_t^p)^2\left[(\lambda-\lambda_0)\cdot 1\{H_t \geq \eta^p\}+\lambda_0\right]}{\kappa\cdot 1\{H_t^p\geq\eta^p\}+H_t^p}.
\end{aligned}
$$

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

$$
\left|q_t^p(k,k')\right| = \frac{H_t^p|k'-k|}{2(H_t^p+\kappa)} < \frac{H_t|k'-k|}{2(H_t+\kappa)} = \left|q_t(k,k')\right|
$$

whenever there is active reallocation. **Q.E.D.**

D Extension: 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.

$$S_t\left(k,k'\right) = \frac{H_t^2\left(k'-k\right)^2}{\kappa_1\left(\varepsilon+\varepsilon'\right)+2\kappa_0+2H_t}, \tag{D.23}$$

$$q_t\left(k,k'\right) = \frac{H_t\left(k'-k\right)}{\kappa_1\left(\varepsilon+\varepsilon'\right)+2\kappa_0+2H_t}. \tag{D.24}$$

The choice of optimal search intensity is given by equation (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

$$\max_{R,q}\left[V_t\left(k+q\right)-e^{-r(T-t+\Delta)}R-V_t\left(k\right)-\chi\left(0,q\right)\right]^{\theta} \times\left[\tilde{V}_t\left(\tilde{k}-q\right)+e^{-r(T-t+\Delta)}R-\tilde{V}_t\left(\tilde{k}\right)\right]^{1-\theta}.$$

The maximized surplus and optimal trade size are given by

$$\tilde{S}_t\left(k,\tilde{k}\right) = \frac{\left[E_t-\tilde{E}_t+2\left(\tilde{H}_t\tilde{k}-H_tk\right)\right]^2}{4\left[H_t+\tilde{H}_t+\kappa_0\right]}, \tag{D.25}$$

$$\tilde{q}_t\left(k,\tilde{k}\right) = \frac{E_t-\tilde{E}_t+2\left(\tilde{H}_t\tilde{k}-H_tk\right)}{2\left[H_t+\tilde{H}_t+\kappa_0\right]}. \tag{D.26}$$

Therefore, the HJB for peripheral traders is

$$r\tilde{V}_t\left(\tilde{k}\right) = \dot{\tilde{V}}_t\left(\tilde{k}\right)+\left(1-\theta\right)\varphi\int\tilde{S}_t\left(k,\tilde{k}\right)dF_t\left(k\right).$$

By matching coefficients we can obtain

$$\begin{aligned}
\dot{\tilde{H}}_t &= r\tilde{H}_t+\frac{\left(1-\theta\right)\varphi\tilde{H}_t^2}{H_t+\tilde{H}_t+\kappa_0}, \text{ with } \tilde{H}_T=0; \\
\dot{\tilde{E}}_t &= r\tilde{E}_t-\left(1-\theta\right)\varphi\tilde{H}_t\frac{E_t-\tilde{E}_t-2H_tK_t}{H_t+\tilde{H}_t+\kappa_0}, \text{ with } \tilde{E}_T=1+i^{RRP}; \\
\dot{\tilde{D}}_t &= r\tilde{D}_t-\left(1-\theta\right)\varphi\int\frac{\left[E_t-\tilde{E}_t-2H_tk\right]^2}{4\left[H_t+\tilde{H}_t+\kappa_0\right]}dF_t\left(k\right), \text{ with } \tilde{D}_t=0.
\end{aligned}$$

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

$$\begin{aligned}
1+\tilde{\rho}_t\left(k,\tilde{k}\right) &= e^{r(T+\Delta-t)}\left[\frac{1-\theta}{2}\left(E_t-\tilde{E}_t-2H_tk\right)+\tilde{E}_t\right] \\
&= e^{r(T+\Delta-t)}\left[\left(1-\theta\right)\left(H_t+\kappa_0\right)\tilde{q}_t\left(k,\tilde{k}\right)+\tilde{E}_t\right]
\end{aligned}$$

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

$$\dot{H}_t = rH_t - a_2 + \frac{1}{4}\frac{H_t^2}{\kappa_1\varepsilon_t + \kappa_0 + H_t}\left[(\lambda - \lambda_0)\varepsilon_t^2 + \lambda_0\right] + \frac{\theta\varphi\vartheta H_t^2}{H_t + \kappa_0}, \quad \text{(D.27)}$$

$$\dot{E}_t = rE_t - a_1 + \frac{K_t}{2}\frac{H_t^2}{\kappa_1\varepsilon_t + \kappa_0 + H_t}\left[(\lambda - \lambda_0)\varepsilon_t^2 + \lambda_0\right] + \theta\varphi\vartheta H_t \frac{E_t - \tilde{E}_t}{H_t + \kappa_0}, \quad \text{(D.28)}$$

$$D_t = rD_t - \frac{1}{4}\frac{H_t^2}{\kappa_1\varepsilon_t + \kappa_0 + H_t}\left[(\lambda - \lambda_0)\varepsilon_t^2 + \lambda_0\right] \int k'^2 dF_t(k') - \theta\varphi\vartheta \frac{\left[E_t - \tilde{E}_t\right]^2}{4\left[H_t + \kappa_0\right]}. \quad \text{(D.29)}$$

It follows that

$$\frac{d\left(E_t - \tilde{E}_t\right)}{dt} = \left(r + \frac{\theta\rho\varphi H_t}{H_t + \kappa_0}\right)\left(E_t - \tilde{E}_t\right) - a_1 + \frac{K_t}{2}\frac{H_t^2}{\kappa_1\varepsilon_t + \kappa_0 + H_t}\left[(\lambda - \lambda_0)\varepsilon_t^2 + \lambda_0\right], \quad \text{(D.30)}$$

and

$$\dot{K}_t = \varphi\vartheta \int \frac{E_t - \tilde{E}_t + 2\left(\tilde{H}_t\tilde{k} - H_t k\right)}{2\left[H_t + \tilde{H}_t + \kappa_0\right]} dF_t(k) = \varphi\vartheta \frac{E_t - \tilde{E}_t - 2H_tK_t}{2\left[H_t + \kappa_0\right]}, \quad \text{(D.31)}$$

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)

$$\begin{aligned} 0 = {} & \left(r + \frac{\lambda}{4} + \theta\varphi\vartheta\right) H^3 + \left[r(2\kappa_0 + \kappa_1) - a_2 + \frac{\kappa_0\lambda}{4} + \theta\varphi\vartheta(\kappa_0 + \kappa_1)\right] H^2 \\ & + \left[r\kappa_0(\kappa_0 + \kappa_1) - a_2(2\kappa_0 + \kappa_1)\right] H - a_2\kappa_0(\kappa_0 + \kappa_1). \end{aligned}$$

Let $A \equiv r + \frac{\lambda}{4} + \theta\varphi\vartheta$, $B \equiv r(2\kappa_0 + \kappa_1) - a_2 + \frac{\kappa_0\lambda}{4} + \theta\varphi\vartheta(\kappa_0 + \kappa_1)$, $C \equiv r\kappa_0(\kappa_0 + \kappa_1) - a_2(2\kappa_0 + \kappa_1)$ and $D \equiv -a_2\kappa_0(\kappa_0 + \kappa_1)$, then the solution to ω_1, ω_2 and ω_3 are given by

$$\begin{aligned} \omega_1 = {} & \frac{-B}{3A} + \sqrt[3]{\frac{BC}{6A^2} - \frac{B^3}{27A^3} - \frac{D}{2A} + \sqrt{\left(\frac{BC}{6A^2} - \frac{B^3}{27A^3} - \frac{D}{2A}\right)^2 + \left(\frac{C}{3A} - \frac{B^2}{9A^2}\right)^3}} \\ & + \sqrt[3]{\frac{BC}{6A^2} - \frac{B^3}{27A^3} - \frac{D}{2A} - \sqrt{\left(\frac{BC}{6A^2} - \frac{B^3}{27A^3} - \frac{D}{2A}\right)^2 + \left(\frac{C}{3A} - \frac{B^2}{9A^2}\right)^3}}, \end{aligned}$$

$$\begin{aligned} \omega_2 = {} & \frac{-B}{3A} + \frac{-1 + \sqrt{3}i}{2}\sqrt[3]{\frac{BC}{6A^2} - \frac{B^3}{27A^3} - \frac{D}{2A} + \sqrt{\left(\frac{BC}{6A^2} - \frac{B^3}{27A^3} - \frac{D}{2A}\right)^2 + \left(\frac{C}{3A} - \frac{B^2}{9A^2}\right)^3}} \\ & + \frac{-1 - \sqrt{3}i}{2}\sqrt[3]{\frac{BC}{6A^2} - \frac{B^3}{27A^3} - \frac{D}{2A} - \sqrt{\left(\frac{BC}{6A^2} - \frac{B^3}{27A^3} - \frac{D}{2A}\right)^2 + \left(\frac{C}{3A} - \frac{B^2}{9A^2}\right)^3}}, \end{aligned}$$

$$\omega_3 = \frac{-B}{3A} + \frac{-1-\sqrt{3}i}{2}\sqrt[3]{\frac{BC}{6A^2} - \frac{B^3}{27A^3} - \frac{D}{2A} + \sqrt{\left(\frac{BC}{6A^2} - \frac{B^3}{27A^3} - \frac{D}{2A}\right)^2 + \left(\frac{C}{3A} - \frac{B^2}{9A^2}\right)^3}}$$
$$+\frac{-1+\sqrt{3}i}{2}\sqrt[3]{\frac{BC}{6A^2} - \frac{B^3}{27A^3} - \frac{D}{2A} - \sqrt{\left(\frac{BC}{6A^2} - \frac{B^3}{27A^3} - \frac{D}{2A}\right)^2 + \left(\frac{C}{3A} - \frac{B^2}{9A^2}\right)^3}}.$$

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

$$\tilde{\mu}_1 \equiv \frac{1}{2r+\frac{\lambda_0}{2}+2\theta\rho\varphi}\left\{-(\kappa_0 r - a_2) - \left[(\kappa_0 r - a_2)^2 + a_2\kappa_0(4r+\lambda_0+4\theta\rho\varphi)\right]^{0.5}\right\},$$
$$\tilde{\mu}_2 \equiv \frac{1}{2r+\frac{\lambda_0}{2}+2\theta\rho\varphi}\left\{-(\kappa_0 r - a_2) + \left[(\kappa_0 r - a_2)^2 + a_2\kappa_0(4r+\lambda_0+4\theta\rho\varphi)\right]^{0.5}\right\},$$
$$\tilde{\tau}_1(H;A,u) \equiv u - \frac{1}{r+\frac{\lambda}{4}+\theta\rho\varphi}\left[\beta_1\log\left(\frac{A-\omega_1}{H-\omega_1}\right) + \beta_2\log\left(\frac{A-\omega_2}{H-\omega_2}\right) + \beta_3\log\left(\frac{A-\omega_3}{H-\omega_3}\right)\right],$$

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 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. **Q.E.D.**

E Extension: Federal Funds Brokerage

In this section we model the brokerage of Federal funds following Lagos & 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:

$$Y_t\left(k, k'\right) = \arg\max_y \left\{y\left[0.5S_t\left(k, k'\right) - y\right]\right\}.$$

Hence the bargaining solution is

$$Y_t\left(k, k'\right) = Y_t\left(k', k\right) = 0.25S_t\left(k, k'\right).$$

The value of the broker, J_t, solves the following HJB equation

$$rJ_t = \dot{J}_t + \alpha \int\int \left[Y_t\left(k, k'\right) + Y_t\left(k', k\right)\right] dF_t\left(k'\right) dF_t(k), \text{ where } J_T = 0.$$

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

$$rV_t(k) = \dot{V}_t(k) + u(k) + \max_{\varepsilon_t \in [0,1]} \int \frac{1}{2} S_t\left(k, k', \varepsilon_t, \varepsilon_t\left(k'\right)\right) m\left(\varepsilon_t, \varepsilon_t\left(k'\right)\right) dF_t\left(k'\right) + \alpha\nu \int \frac{1}{4} S_t\left(k, k', 0, 0\right) dF_t\left(k'\right)$$

With quadratic utility function, we guess and verify $V_t(k) = -H_t k^2 + E_t k + D_t$, the solution is

$$\begin{aligned} rV_t(k) &= \dot{V}_t(k) + u(k) + \frac{1}{2}\frac{(H_t)^2}{\kappa(\varepsilon+\varepsilon_t)+2H_t}\left[(\lambda-\lambda_0)\varepsilon\varepsilon_t+\lambda_0\right]\int\left(k'-k\right)^2 dF_t\left(k'\right) \\ &+\frac{\alpha\nu}{8}H_t\int\left(k'-k\right)^2 dF_t\left(k'\right) \end{aligned}$$

By matching coefficients we obtain

$$\left(r+\frac{\alpha\nu}{8}\right)H_t = \dot{H}_t + a_2 - \frac{1}{4}\frac{H_t^2}{\kappa\varepsilon_t+H_t}\left[(\lambda-\lambda_0)\varepsilon_t^2+\lambda_0\right], \tag{E.32}$$

thus $\alpha\nu$ changes the discount rate to the banks. The surplus function is

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

and the broker's HJB is

$$\begin{aligned} rJ_t &= \dot{J}_t + \frac{\alpha}{2}\int\int S_t\left(k,k',0\right)dF_t\left(k'\right)dF_t(k) \\ &= \dot{J}_t + \frac{\alpha}{4}H_t\left[\int k^2 dF_t(k) - K^2\right] \end{aligned} \tag{E.33}$$

The solution is

$$J_0(\nu) = \frac{\alpha}{4}\int_0^T e^{-rt}H_t\left[\int k^2 dF_t(k) - K^2\right]dt, \tag{E.34}$$

where

$$\begin{aligned} &\int k^2 dF_t(k) \\ =\ &\int k^2 dF_0(k)\exp\left\{-\int_0^t m\left(\varepsilon_z,\varepsilon_z\right)\frac{H_z\left(H_z+2\kappa\varepsilon_z\right)}{2\left(H_z+\kappa\varepsilon_z\right)^2}dz - \frac{\alpha\nu}{2}t\right\} \\ &+K^2\int_0^t \exp\left\{-\int_z^t m\left(\varepsilon_s,\varepsilon_s\right)\frac{H_s\left(H_s+2\kappa\varepsilon_s\right)}{2\left(H_s+\kappa\varepsilon_s\right)^2}ds - \frac{\alpha\nu}{2}(t-z)\right\}\left[m\left(\varepsilon_z,\varepsilon_z\right)\frac{H_z\left(H_z+2\kappa\varepsilon_z\right)}{2\left(H_z+\kappa\varepsilon_z\right)^2}+\frac{\alpha\nu}{2}\right]dz \\ =\ &K^2 + \left[\int k^2 dF_0(k) - K^2\right]\exp\left\{-\int_0^t m\left(\varepsilon_z,\varepsilon_z\right)\frac{H_z\left(H_z+2\kappa\varepsilon_z\right)}{2\left(H_z+\kappa\varepsilon_z\right)^2}dz - \frac{\alpha\nu}{2}t\right\} \end{aligned}$$

Thus the solution to $J_0(\alpha)$ can be written as

$$J_0(\nu) = \frac{\alpha}{4}\left[\int k^2 dF_0(k) - K^2\right]\int_0^T e^{-rt}H_t\exp\left\{-\int_0^t m\left(\varepsilon_z,\varepsilon_z\right)\frac{H_z\left(H_z+2\kappa\varepsilon_z\right)}{2\left(H_z+\kappa\varepsilon_z\right)^2}dz - \frac{\alpha\nu}{2}t\right\}dt.$$

Thus the equilibrium matchmaking is

$$\psi = \frac{\alpha}{4}\left[\int k^2 dF_0(k) - K^2\right]\int_0^T e^{-rt}H_t\exp\left\{-\int_0^t m\left(\varepsilon_z,\varepsilon_z\right)\frac{H_z\left(H_z+2\kappa\varepsilon_z\right)}{2\left(H_z+\kappa\varepsilon_z\right)^2}dz - \frac{\alpha\nu}{2}t\right\}dt. \tag{E.35}$$

The following proposition characterizes the comparative statics of the equilibrium measure of brokers with respect to policy and technology parameters.

Proposition 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

$$\frac{\dot{M}_t}{M_t} = -\frac{3}{8}\alpha\nu - \frac{a_2}{H_t} - \frac{m(\varepsilon_t, \varepsilon_t)}{4} \frac{H_t}{H_t + \kappa\varepsilon_t} \left(\frac{2\kappa\varepsilon_t}{H_t + \kappa\varepsilon_t} + 1 \right) < 0,$$

and

$$\frac{\partial}{\partial H_t} \left(\frac{\dot{M}_t}{M_t} \right) = \frac{a_2}{H_t} - \frac{m(\varepsilon_t, \varepsilon_t)}{4} \frac{\kappa\varepsilon_t}{(H_t + \kappa\varepsilon_t)^2} \left(\frac{4\kappa\varepsilon_t}{H_t + \kappa\varepsilon_t} - 1 \right).$$

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

$$\frac{\partial J_0(\nu)}{\partial A_2} = \int_0^T \frac{\partial J_0(\nu)}{\partial M_t} \frac{\partial M_t}{\partial A_2} dt > 0.$$

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.**

F Extension: 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

$$\begin{aligned} rV_t(k) &= \dot{V}_t(k) + u(k) + \max_{\varepsilon \in [0,1]} \int \frac{1}{2} S_t(k, k', \varepsilon, \varepsilon_t(k')) \, m(\varepsilon, \varepsilon_t(k')) \, dF_t(k') \qquad \text{(F.36)} \\ &\quad + \zeta \int [V_t(k + w) - V_t(k)] \, dG(w) + \mu \frac{\partial}{\partial k} V_t(k) + \frac{\sigma_C^2}{2} \frac{\partial^2}{\partial k^2} V_t(k). \end{aligned}$$

Given $\{F_t\}$, the value function in an equilibrium is given by

$$V_t(k) = -H_t k^2 + E_t k + D_t, \tag{F.37}$$

where H_t, E_t and D_t are given by the solutions to the following initial-value ODE problems

$$\dot{H}_t = rH_t - a_2 + \frac{1}{4}\frac{H_t^2\left[(\lambda-\lambda_0)\varepsilon_t^2+\lambda_0\right]}{\kappa\varepsilon_t + H_t}, \tag{F.38}$$

$$\dot{E}_t = rE_t - a_1 + \frac{K_t}{2}\frac{H_t^2\left[(\lambda-\lambda_0)\varepsilon_t^2+\lambda_0\right]}{\kappa\varepsilon_t + H_t} + 2\mu H_t, \tag{F.39}$$

$$\dot{D}_t = rD_t - \frac{1}{4}\frac{H_t^2\left[(\lambda-\lambda_0)\varepsilon_t^2+\lambda_0\right]}{\kappa\varepsilon_t + H_t}\int k'^2 dF_t(k') + \left(\zeta\sigma_L^2 + \sigma_C^2\right)H_t - \mu E_t, \tag{F.40}$$

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

$$\varepsilon_t(k) = \begin{cases} 1, & \text{if } H_t \geq \eta; \\ 0, & \text{otherwise.} \end{cases} \tag{F.41}$$

The Federal funds purchased $q_t(k,k')$ and the Federal funds rate $\rho_t(k,k')$ are given by

$$q_t(k,k') = \frac{H_t(k'-k)}{2(\kappa\varepsilon_t + H_t)}, \tag{F.42}$$

$$\rho_t(k,k') = e^{r(T+\Delta-t)}\left[E_t - H_t(k+k')\right]. \tag{F.43}$$

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 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.

G Extension: 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

$$p_L\left[V_t\left(k+q\right)-p_Be^{-r(T+\Delta-t)}R\right]-p_LV_t\left(k\right)-\chi\left(\varepsilon,q\right).$$

The lender's surplus is given by

$$p_L\left[V_t\left(k'-q\right)+p_Be^{-r(T+\Delta-t)}R\right]-p_LV_t\left(k'\right)-\chi\left(\varepsilon',-q\right).$$

The solution to Nash bargaining problem becomes

$$q_t\left(k,k',\varepsilon,\varepsilon'\right)=\frac{H_t\left(k'-k\right)}{\kappa/p_L\cdot\left(\varepsilon+\varepsilon'\right)+2H_t}, \tag{G.44}$$

$$R_t\left(k,k',\varepsilon,\varepsilon'\right)=\frac{e^{r(T+\Delta-t)}}{p_B}\left[E_t-H_t\left(k+k'\right)-\frac{\kappa\left(\varepsilon-\varepsilon'\right)}{2p_L}q_t\left(k,k',\varepsilon,\varepsilon'\right)\right]q_t\left(k,k',\varepsilon,\varepsilon'\right), \tag{G.45}$$

$$\rho\left(k,k',\tau\right)=\frac{R\left(k,k',\tau\right)}{q\left(k,k,\tau\right)}=\frac{e^{r(T+\Delta-t)}}{p_B}\left[E_t-H_t\left(k+k'\right)-\frac{\kappa\left(\varepsilon-\varepsilon'\right)}{2p_L}q_t\left(k,k',\varepsilon,\varepsilon'\right)\right], \tag{G.46}$$

$$S_t\left(k,k',\varepsilon,\varepsilon'\right)=\frac{p_L\left[H_t\left(k'-k\right)\right]^2}{\kappa/p_L\cdot\left(\varepsilon+\varepsilon'\right)+2H_t}. \tag{G.47}$$

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 $V_t\left(k\right)=-H_tk^2+E_tk+D_t$, where

$$\dot{H}_t = rH_t-a_2+\frac{p_L}{4}\frac{H_t^2}{\kappa/p_L\cdot\varepsilon_t+H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2+\lambda_0\right],\ \text{where } H_T=A_2; \tag{G.48}$$

$$\dot{E}_t = rE_t-a_1+\frac{Kp_L}{2}\frac{H_t^2}{\kappa/p_L\cdot\varepsilon_t+H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2+\lambda_0\right],\ \text{where } E_T=A_1; \tag{G.49}$$

$$\dot{D}_t = rD_t-\frac{p_L}{4}\frac{H_t^2}{\kappa/p_L\cdot\varepsilon_t+H_t}\left[\left(\lambda-\lambda_0\right)\varepsilon_t^2+\lambda_0\right]\int k'^2dF_t\left(k'\right),\ \text{where } D_T=0. \tag{G.50}$$

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.

H 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 12 and solve the paths of E_t and K_t via the ODEs (D.30) and (D.31).

2. Given the path of ε_t, simulate a Poisson process for bilateral meetings up to time T via a thinning algorithm:[24]

   (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.

   (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 $U_{(1)} < U_{(2)} < ... < U_{\left(\hat{M}\right)}$.

   (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.

[24] See Sigman (2007) for a detailed description and proof of the thinning algorithm.

(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.[25] For each quarter, we simulate the model for $S = 2,000$ times.

[25] See Adda & Cooper (2003) for the reference on simulated method of moments.