NBER WORKING PAPER SERIES

LIQUIDITY REGULATION AND UNINTENDED FINANCIAL TRANSFORMATION IN CHINA

Kinda Cheryl Hachem
Zheng Michael Song

Working Paper 21880
http://www.nber.org/papers/w21880

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2016

An earlier draft of this paper circulated under the title "The Rise of China's Shadow Banking System." We thank Anil Kashyap for several helpful conversations. We also thank Jeff Campbell, Doug Diamond, Chang-Tai Hsieh, John Huizinga, Mark Kruger, Rich Rosen, Ben Sawatzky, Martin Schneider, and Harald Uhlig. Special thanks as well to Mingkang Liu and the other bankers and regulators we interviewed. We are also grateful to seminar and conference participants at Minnesota Carlson, CUHK, Queen's University, Bank of Canada, Chicago Booth, Nankai, HKU, Tsinghua, Fudan University, FRB Minneapolis, NUS, SED 2015, NBER Summer Institute 2015 (Monetary Economics Workshop), and FRB Chicago. Financial support from Chicago Booth is gratefully acknowledged. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Liquidity Regulation and Unintended Financial Transformation in China
Kinda Cheryl Hachem and Zheng Michael Song
NBER Working Paper No. 21880
January 2016
JEL No. E44,G28

ABSTRACT

China began setting higher liquidity standards for its banking system in the late 2000s. This proved counter-productive: the interbank market became tighter and more volatile and credit soared. Our paper explains what happened and why. We first document that shadow banking emerged among China's small and medium-sized banks to evade the higher liquidity standards. We then argue that shadow banking poached deposits from big commercial banks and, in response, big banks used their interbank market power to try and undermine the shadow banks. A calibration of our model generates a quantitatively important credit boom and connects an otherwise disperse set of facts about China's rising financial system.

Kinda Cheryl Hachem
University of Chicago
Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
kinda.hachem@chicagobooth.edu

Zheng Michael Song
Department of Economics
Chinese University of Hong Kong
Shatin, N.T., Hong Kong
zheng.michael.song@gmail.com

1 Introduction

China began setting higher liquidity standards for its banking system in the late 2000s. The anticipated consequence was a more liquid interbank market and a moderation in credit. Neither of these things happened. Instead, the interbank market became tighter and more volatile and credit soared. Was China's experience a coincidence or does it bring to light some deeper economic forces? Answering this question is important as we embark on a new era of banking regulation after the 2007-09 financial crisis.

Our paper shows that China's experience was not a coincidence. We start by establishing a new and comprehensive set of facts about the Chinese financial system. We then show that liquidity regulation has been made counter-productive by two economic forces: regulatory arbitrage by small banks and a response by big banks using their interbank market power. Quantitatively, the combination of these two forces explains a sizeable amount of the data. Our analysis highlights the importance of using micro-founded models to predict the effect of regulation. It also shows that interbank market structure can re-orient the monetary transmission mechanism, particularly in transitioning economies.

To explain the forces in our paper and how they interact, let us first explain the co-existence of big and small banks in China's financial system. One job of any financial system is to connect savings with investment opportunities. In a well-functioning system, intermediaries identify suitable borrowers and attract enough savings to finance these borrowers by offering savers sufficiently high interest rates. China's government interferes with this process in two ways. First, it regulates interest rates. The main interest rate regulation has traditionally been an uncompetitive ceiling on deposit rates which favors banks with deep/entrenched retail networks. State-owned banks have such networks so they have been kept big while other banks have been kept small. Second, China's government interferes with lending by imposing loan limits on all banks and dissuading state-owned banks from financing the private sector.[1] Non-state-owned banks thus serve the private sector but are limited by their small size.

Until recently, some of these government distortions were smoothed over by the interbank market: small banks channeled almost all of their existing deposits into non-financial loans then borrowed from big banks when in need of extra/emergency liquidity. The big banks, flush with deposits, were willing to lend to small banks at an appropriate rate rather than make additional loans to the tepid state sector or the politically thorny private sector.

[1] For more on the emergence of China's private sector, see Hsieh and Klenow (2009), Song et al (2011), Brandt et al (2012), Lardy (2014), Hsieh and Song (2015), and the references therein.

Things changed around 2008 when the government began enforcing a loan-to-deposit cap which forbids banks from lending more than 75% of their deposits to non-financial firms. Enforcement was complemented by a large increase in reserve requirements, making the 75% cap akin to a liquidity standard. Big banks had loan-to-deposit ratios well below 75% so the stricter rules were essentially aimed at small banks. The first part of our paper shows that small banks responded by engaging in regulatory arbitrage. In particular, small banks began offering a new savings instrument called a "wealth management product" (or WMP for short) that could be used to circumvent the regulation. As long as the WMP does not come with an explicit principal guarantee from the issuing bank, it does not need to be reported on the bank's balance sheet. Instead, the savings attracted by the WMP are funneled into a trust company which makes the loans that small banks cannot make without violating the 75% loan-to-deposit cap.[2]

The bank-trust cooperation just described constitutes shadow banking: it achieves the same type of credit intermediation as a regular bank without appearing on a regulated balance sheet. It also achieves the same type of maturity transformation as a regular bank, with long-term assets financed by short-term liabilities (e.g., the average trust loan matures in about two years while the average WMP matures in three months). As per Diamond and Dybvig (1983), this mismatch creates a liquidity service for savers but is highly runnable without government insurance.[3]

Among industry analysts, there is a sense that stricter liquidity rules have contributed to shadow banking in China. However, the discussion lacks identification and exact mechanisms have not been clearly established. Our paper exploits cross-sectional differences in WMP issuance along with time-varying WMP characteristics to show that stricter liquidity rules were the trigger for China's shadow banking system. We then present a simple model which is able to generate shadow banking as an endogenous response to regulation, corroborating the empirical evidence.

The second part of our paper shows how to get from the shadow banking activities of small banks to tighter interbank conditions and higher credit. To build some intuition about

[2]That smaller banks are the driving force behind regulatory arbitrage in China stands in sharp contrast to other regions. In the U.S. and Europe, big banks are generally seen as the main drivers. See, for example, Cetorelli and Peristiani (2012) and Acharya et al (2013).

[3]In this regard, there is a notable similarity between unguaranteed WMPs and the asset-backed commercial paper vehicles that collapsed during the 2007-09 financial crisis. Another similarity is the use of implicit guarantees and/or back-up credit lines by the sponsoring bank to allay investor concerns about the riskiness of off-balance-sheet products. As a result, the products are only off-balance-sheet for accounting purposes. For more on the 2007-09 crisis, see Brunnermeier (2009), Gorton and Metrick (2012), Covitz et al (2013), Kacperczyk and Schnabl (2013), Krishnamurthy et al (2014), and the references therein.

the necessary ingredients, we start with a Walrasian interbank market where banks hit by high liquidity shocks can borrow from banks hit by low liquidity shocks. These shocks stem from the maturity mismatch between bank assets and liabilities. In the Walrasian environment, the interest rate that clears the interbank market is highest when there is no liquidity regulation. This result is the market mechanism at work. A low interbank rate eliminates the precautionary savings motive of each individual bank, resulting in an insufficient amount of aggregate liquidity. This cannot be an equilibrium so the interbank rate must be high in the absence of government intervention. If a liquidity minimum is introduced, banks will have to reduce lending to meet the minimum and, with more liquidity in the system, the interbank market will clear at a lower interest rate.[4]

A single yet powerful modification overturns these predictions to explain the broader financial trends in China: introducing a big player on the interbank market. Each individual small bank remains an interbank price-taker while the big bank is effectively a price-setter. Liquidity regulation still leads to shadow banking by small banks but now there is competition between shadow banking and the big bank: by virtue of being off-balance-sheet, WMPs can breach the uncompetitive ceiling on deposit rates and poach a lot of funds from the big bank. As part of its equilibrium response, the big bank changes its interbank behavior. Given the maturity mismatch between WMPs and trust loans, the interbank market remains an important source of emergency liquidity for small banks. Therefore, by cutting such liquidity, the big bank can make small banks less aggressive in WMP issuance and lessen the extent to which small banks poach funds. This strategy makes the interbank market tighter and more volatile, consistent with the data. In the interbank repo market, for example, the weighted average interest rate increased by 50 basis points between 2007 and 2014 despite increasing monetary injections by the central bank. The maximum daily rate also increased by 150 basis points after 2007, reaching an unprecedented 11.6% in mid-2013. We study the mid-2013 event using high frequency data and show that big banks as a group are indeed manipulating the interbank market against small banks.

To recap, we argue that enforcement of the 75% loan-to-deposit cap pushed small banks into shadow banking which then pushed big banks to tighten the interbank market. Our paper thus provides a novel explanation for why China's interbank markets became more volatile at a time when regulatory policies were designed to increase bank liquidity. Our paper also speaks to the large credit boom that has taken place in China. First, the reallocation of savings from deposits at the big banks to higher-return WMPs at the small banks increases

[4] See Farhi et al (2009) for a different environment in which a liquidity floor decreases interest rates.

total credit because the small banks (and their trusts) typically lend more per unit of savings than the big banks. Second, the strategic interbank response of the big banks increases credit through traditional lending: rather than sitting idle on the liquidity that they intend to withhold from the interbank market, the big banks lend more to non-financial borrowers. Stricter liquidity rules can thus lead to more credit, not less, when one takes into account interactions between heterogeneous banks.[5]

A calibration of our model generates one-third of the increase in China's aggregate credit-to-savings ratio between 2007 and 2014 and over one-half of the increase in interbank rates over the same period. These are sizeable magnitudes which challenge the conventional wisdom that most of China's credit boom was caused by a bank-funded fiscal stimulus package undertaken by the central government in 2009 and 2010. A money multiplier calculation shows that stimulus alone explains roughly the same fraction of the boom as the mechanisms in our model.

The rest of the paper proceeds as follows. Section 2 describes the basic features of China's banking system, Section 3 presents the new empirical facts, Section 4 builds a model to rationalize the facts, Section 5 calibrates the model, and Section 6 concludes. All proofs are in the appendix.

2 Institutional Background

We begin with the institutional features that surround the rise of shadow banking in China. We first describe the main players in the regulated banking sector (Section 2.1) and the banking regulations they face (Section 2.2). We then document how these regulations are being circumvented (Section 2.3) and how large the resulting shadow sector has become (Section 2.4).

2.1 Traditional Banking in China

Until the late 1970s, China had a Soviet-style financial system where the central bank was the only bank. The Chinese government moved away from this system in the late 1970s and early 1980s by establishing four state-owned commercial banks (the Big Four). Market-oriented reforms initiated in the 1990s then led to two additional changes.

[5] In this regard, our paper also contributes to the literature on the industrial organization of banking (e.g., Corbae and D'Erasmo (2013) and the references therein).

First, the Big Four became much more profit-driven. All went through a major restructuring in the mid-2000s and are now publicly listed. Controlling interest is still held by the government[6] but the government now only limits (i) how intensely the Big Four compete with each other and (ii) how much they lend to the private sector. Previously, the government was involved in almost all aspects of bank decision-making. The effect of the reforms has been striking. The average non-performing loan ratio of the Big Four fell from 30% in 2000 to roughly 2% in recent years. Combined profits also grew 19% annually from 2007 to 2014 to reach an unprecedented USD 184 billion in 2014. Individually, the banks in China's Big Four now constitute the first, second, fourth, and seventh largest banks in the world as measured by total assets.[7]

The second notable change was entry of small and medium-sized commercial banks. China now has twelve joint-stock commercial banks (JSCBs) operating nationally and over one hundred city banks operating in specific regions. Many rural banks have also emerged. The JSCBs are typically larger than the city and rural banks but all of these banks are individually small when compared to the Big Four. For example, average deposits for the JSCBs were only 17% of average deposits for the Big Four in 2013. As a group though, small and medium-sized banks have chipped away at the Big Four's deposit share. In 1995, the Big Four held 80% of deposits in China. By 2005, they held 60%. The Big Four now account for roughly 50% of traditional deposits.[8]

2.2 Banking Regulations

China's banks are regulated by two agencies: the China Banking Regulatory Commission (CBRC) and the People's Bank of China (central bank). CBRC was established in 2003 to take over banking supervision from the central bank. One part of China's regulatory environment, capital regulation, has been shaped by the international Basel Accords.[9] However, two historically important regulations – a ceiling on bank deposit rates and a cap on bank loan-to-deposit ratios – did not stem from similar international standards.

[6] Logistically, the Ministry of Finance and Huijin (a government-owned investment company) retain controlling interests while the Organization Department of the Chinese Communist Party appoints top executives. Similar features are found in big state-owned firms in the industrial sector (Hsieh and Song (2015)).

[7] http://www.relbanks.com/worlds-top-banks/assets

[8] Big state-owned firms in the industrial sector did not experience a similar post-restructuring decline in market share.

[9] After CBRC was established, it introduced an 8% minimum capital adequacy ratio as per Basel I. The higher requirements of Basel III are currently being phased in. CBRC will require a minimum capital adequacy ratio of 11.5% for systemically important banks and 10.5% for all other banks by the end of 2018. The requirements were 9.5% and 8.5% respectively at the end of 2013. For comparison, the Big Four had an average capital ratio of 12.7% in 2013 while the average across all Chinese banks in Bankscope was 14.7%.

China has a long history of regulating deposit rates. Prior to 2004, deposit rates were simply set by the central bank. In 2004, downward flexibility was introduced and deposit rates were allowed to fall below the central bank's benchmark rate. All banks stayed at the benchmark, revealing it as a binding ceiling. Some upward flexibility was then introduced in 2012 when the central bank allowed deposit rates of up to 1.1 times the benchmark rate. Almost all banks for which we have systematic data responded by setting the maximum allowable deposit rate so, once again, the central bank's ceiling proved binding.[10]

The next historically important regulation, a 75% loan-to-deposit cap, was written into China's Law on Commercial Banks in 1995. Enforcement of the cap was loose until 2008 when CBRC moved to rein in loan growth at small and medium-sized banks. At first, CBRC policed end-of-year loan-to-deposit ratios. It switched to end-of-quarter ratios in late 2009, end-of-month ratios in late 2010, and average daily ratios in mid-2011. Stricter loan-to-deposit rules were also complemented by a rapid increase in the reserve requirements set by the central bank. Official requirements were 7.5% in 2005. There was a modest increase to 9.5% by early 2007 then a rapid climb to 15.5% by February 2010. Reserve requirements were subsequently increased twelve times to reach 21.5% by December 2011.[11]

2.3 Bank-Trust Cooperation as Regulatory Arbitrage

With the main banking regulations in hand, we now discuss how regulation has been circumvented. Wealth management products (WMPs) are the centerpiece of regulatory arbitrage in China. A WMP is a savings instrument that is typically sold at bank counters. WMPs have two features which help get around the regulations discussed above. First, WMP returns are not subject to a deposit rate ceiling. Figure 1 plots data on annualized WMP returns. The spread relative to the one-year deposit rate has averaged 1 percentage point since 2008 and nearly 2 percentage points since 2012.[12] Second, WMPs do not have to be principal-

[10] In late 2015, China announced that the deposit rate ceiling would be removed. The response of deposit rates to this announcement has been modest. This is consistent with (i) most interest-sensitive savings having already migrated to the high-yielding WMPs discussed in Section 2.3 and/or (ii) on-balance-sheet deposits still entailing other regulatory costs for banks.

[11] We make two comments here. First, the increasing frequency of CBRC's loan-to-deposit checks between 2009 and 2011 and the complementary hike in reserve requirements may have been responses to other government policies (e.g., a bank-funded stimulus package undertaken by the central government in 2009 and 2010). Section 3.2 will show that the stimulus alone was not large enough to generate a credit boom the size of China's so we are ultimately interested in how much of the boom can be explained by changes in bank regulation, controlling for all other government policies. The model we build in Section 4 and its calibration will help produce this counterfactual. Second, although China lifted the official loan-to-deposit component of its liquidity rules in late 2015, reserve requirements remain high and loan-to-deposit restrictions can still technically be imposed via loan quotas.

[12] To date, almost all WMPs have delivered above or equal to their promised returns.

guaranteed by the issuing bank. Without a guarantee, the WMP and the assets it invests in are not consolidated into the bank's balance sheet and thus not subject to loan-to-deposit rules (or capital requirements). According to CBRC, non-guaranteed products were 70% of total WMP issuance in 2012 and 65% of total WMP issuance in 2013.

Where do the funds from unconsolidated WMPs end up? Figure 2 shows the potential channels. Stock, bond, and money markets are all investment options. However, at least three pieces of evidence suggest that the key recipients of non-guaranteed WMPs are lightly-regulated financial institutions called trust companies. First, there has been a near lockstep evolution of trust company assets under management and WMPs outstanding (Figure 3). Second, the funding for roughly 70% of trust assets comes from money that has already been pooled together by other institutions, sometimes referred to as money raised through single trust products (Figure 4). This is remarkably close to the proportion of WMPs that are not guaranteed. Third, trust companies have responded to recent attempts at WMP regulation. In August 2010, for example, CBRC announced that WMPs could invest at most 30% in trust loans. The composition of trust assets then changed from 63% loans at the end of 2010Q2 to 42% loans by the end of 2011Q3.[13]

Another example comes in March 2013 when CBRC went even further and announced that WMPs could invest at most 35% in non-standard debt assets (e.g., all trust assets). Banks and trusts responded by developing the counterpart business illustrated in Figure 5. In short, money is channeled from WMPs to trusts in two individually compliant steps. The WMP issuer first places WMP money in another bank (or bank-affiliated off-balance-sheet vehicle). The WMP's return is tied to interest earned on this placement so the WMP is said to be backed by interest rate products, not trust assets. However, trust companies appear in the next step. In particular, they issue beneficiary rights to the recipient of the placement who then uses the cash flows from those rights to pay the placement interest.[14] We will see in Section 3.1 that trust beneficiary rights became popular exactly when CBRC began cracking down on direct bank-trust cooperation.

Whether direct or indirect, cooperation between banks and trust companies is important for at least two reasons. First, it allows banks to make loans that might have otherwise violated banking regulations. Second, it involves a strong maturity mismatch. The mismatch

[13] Based on data from the China Trustee Association.

[14] The recipient of the placement can acquire trust beneficiary rights either as an investment receivable or through an "offline" reverse repo. Offline transactions are ones which do not go through the China Foreign Exchange Trade System.

can be gleaned by returning to Figures 1 and 4. Figure 1 shows that WMPs are predominantly short-term products. The median maturity has been around 3 months since 2008 and roughly 25% of WMPs have a maturity of 1 month or less. In contrast, Figure 4 shows that trust companies hold the majority of their assets as loans and long-term investments.[15] Further support for the long-term nature of trust company assets comes from the fact that trusts issued products with an average maturity of 1.7 years when trying to pool money on their own during the first half of 2013.[16]

2.4 Measuring the Shadow Sector

The Financial Stability Board defines shadow banking as "credit intermediation [that] takes place in an environment where prudential regulatory standards ... are applied to a materially lesser or different degree than is the case for regular banks engaged in similar activities" (FSB (2011)). The cooperation between banks and trusts discussed in Section 2.3 satisfies this definition. First, it involves maturity transformation and thus constitutes banking in the sense of Diamond and Dybvig (1983). Second, it is funded by non-guaranteed WMPs which are booked off-balance-sheet and away from regulatory standards. We can therefore use non-guaranteed WMPs to get a conservative estimate of shadow banking in China. WMPs outstanding ballooned from 2% of GDP in 2007 to nearly 25% of GDP in 2014 (Figure 3). Also recall that roughly two-thirds of WMP issuance in 2012 and 2013 was non-guaranteed (CBRC). We thus estimate that China's shadow banking system grew from a negligible fraction of GDP in 2007 to 16% of GDP in 2014.

To get a broader estimate of shadow banking, one can use the widely-cited data on total social financing constructed by China's National Bureau of Statistics.[17] Social financing includes bank loans, corporate bonds, equity, and other financing not accounted for by traditional channels. Roughly one-third of other financing takes the form of undiscounted banker's acceptances.[18] Removing these acceptances then leaves the most shadowy part of other financing, namely loans by trust companies and entrusted firm-to-firm loans. It is an open question how much entrusted lending also involves trust companies so we group trust and entrusted loans into one measure of shadow banking.[19] By this measure, shadow banking

[15] The sectorial composition of trust company assets has become more even over time, with infrastructure and real estate projects losing ground to industrial and commercial enterprises.

[16] Annual Report of the Trust Industry in China (2013).

[17] See, for example, Elliott et al (2015).

[18] A banker's acceptance is basically a guarantee by a bank on behalf of a depositor. More precisely, the bank guarantees that the depositor will repay a third-party at a later date.

[19] Allen et al (2015) study entrusted loans made by publicly traded firms. These firms are required to

grew from 5% of GDP in 2007 to 24% of GDP in 2014. Notice that our conservative estimate of shadow banking based only on bank-trust cooperation still accounts for a sizeable amount of the broader measure.

3 Empirical Facts

This section establishes the core empirical facts that motivate our paper. Section 3.1 shows that loan-to-deposit rules triggered shadow banking among China's small and medium-sized banks (henceforth SMBs). Section 3.2 documents an increase in total credit and shows that China's four biggest banks (the Big Four) have become more aggressive in traditional lending. Section 3.3 shows that the Big Four are also manipulating interbank markets. Our primary dataset is the Wind Financial Terminal which provides information about individual wealth management products. It also provides some information about interbank conditions. In cases where Wind is insufficient, we collect data from bank annual reports, regulatory agencies, and financial association websites.

3.1 Loan-to-Deposit Ratio as Regulatory Trigger

The raw loan-to-deposit ratio across all commercial banks averaged 67% between 2007 and 2013 so the 75% cap described in Section 2 appears slack at the aggregate level. A different story emerges from the cross-section. Figure 6 plots the loan-to-deposit ratios of the Big Four and the joint-stock commercial banks (JSCBs).[20] As a group, the JSCBs are just satisfying the 75% cap, averaging an end-of-year loan-to-deposit ratio of 74% between 2007 and 2013. That the 75% cap is a binding constraint for the JSCBs is evidenced by the fact that these banks have a noticeably higher loan-to-deposit ratio when the ratio is calculated using average balances during the year rather than end-of-year balances. It is not until CBRC increases its monitoring frequency (see Section 2.2) that the difference between the end-of-year and average balance ratios for JSCBs begins to disappear. In contrast, the Big Four are not constrained by the 75% cap: their loan-to-deposit ratio has been comfortably below 75% for the past decade and there is virtually no difference between their average balance and end-of-year ratios.[21] We exploit this cross-sectional difference in Subsection 3.1.1. We will

disclose the loans. The authors find that public firms accounted for 10% of the total amount of entrusted loans reported by the central bank in 2013.

[20] Historical balance sheet data for city and rural banks is spotty (particularly when it comes to average daily balances) so these banks are excluded from Figure 6.

[21] A common story is that the government uses individual loan quotas to impose even stricter limits on big banks. In practice though, quotas are negotiable, particularly for the Big Four who have more bargaining power than SMBs.

then present a case study in Subsection 3.1.2 which connects changes in WMP characteristics with changes in CBRC's monitoring frequency to provide some time-varying evidence on the role of loan-to-deposit rules.

3.1.1 Big Four versus Small and Medium-Sized Banks

Heterogeneity in the bindingness of the 75% cap suggests a natural test: if enforcement of the cap did indeed trigger shadow banking, then we should see small and medium-sized banks moving much more heavily into WMPs (and in particular off-balance-sheet WMPs) than the Big Four. We should also see much higher holdings of trust beneficiary rights by SMBs once CBRC restricts direct cooperation between banks and trusts. We confirm these predictions here.

Between 2008 and 2014, SMBs accounted for 73% of all new WMP batches. The SMBs are thus disproportionately more involved in WMP issuance than the Big Four. The SMBs are also disproportionately more involved in non-guaranteed WMPs. Between 2008 and 2014, SMBs issued 57% of their WMP batches without a guarantee while the Big Four issued 46% of their WMP batches without a guarantee. The gap widens to 62% for SMBs versus 43% for the Big Four in the second half of our sample. These estimates are based on product counts since Wind does not yet have complete data on the total funds raised by each product. However, using data from CBRC and the annual reports of the Big Four, we estimate that SMBs accounted for roughly 64% of non-guaranteed WMP balances outstanding at the end of 2013.[22]

Turning to trust beneficiary rights (TBRs), Figure 7 shows a dramatic rise in TBR holdings among joint-stock banks when CBRC cracked down on bank-trust cooperation in 2013. There was no similar rise in TBR holdings among the Big Four. We can also see from Figure 7 that the joint-stock banks did not sacrifice loans in order to hold TBRs. Instead, they substituted away from balances held at banks and other financial institutions (dashed black line). The effect of this substitution is visible in the blue bars: both the joint-stock banks and the Big Four have experienced decreases in the balances they owe banks. However, unlike the Big Four, the joint-stock banks have attracted sufficiently more balances from non-bank financial institutions (red bars). Off-balance-sheet vehicles that hold unguaranteed WMPs would be classified as non-bank financial institutions. Effectively, the

[22] The entire WMP balance in Bank of China's annual report is described as an unconsolidated balance yet the micro data includes several guaranteed batches for this bank that would not have matured by the end of 2013. We therefore remove Bank of China and rescale the other banks in the Big Four to back out our 64% estimate for SMBs.

off-balance-sheet vehicle of one joint-stock bank places money with another joint-stock bank. The other joint-stock bank then uses returns from its TBR holdings to pay interest on the placement. This is exactly the counterpart business discussed in Section 2.3. In principle, even more counterpart business could be occurring between the vehicles themselves (e.g., vehicles can hold TBRs and placements can occur between vehicles).

We have now documented that shadow banking activities are dominated by SMBs. Granger causality tests bolster this result. In particular, we find that WMP issuance by SMBs causes WMP issuance by big banks (Table 1). The reverse is not true at any reasonable level of significance so the impetus for WMPs is indeed coming from small and medium-sized banks. The intuition goes back to the nature of China's banking regulations. Recall from Section 2.2 that China has historically had a binding ceiling on deposit rates. Such a ceiling stifles deposit rate competition and favors banks with deeper and better-established retail networks (i.e., the Big Four). Also recall that China tightened loan-to-deposit rules just as SMB lending was picking up. Unable to comply with the tighter loan-to-deposit rules by attracting more deposits and unwilling to forgo profitable lending opportunities, SMBs had the most to gain from shadow banking.

In principle, SMBs could also be using off-balance-sheet WMPs to skirt capital requirements. However, data from Bankscope suggests that the average SMB held more than the minimum capital requirement even before CBRC adopted the Basel framework in 2004. This is consistent with our discussion. In principle, banks should only want to skirt capital requirements that force them to switch from cheap funding (deposits) to more expensive funding (capital). However, precisely because cheap deposits are difficult for the average SMB to attract, it makes sense that SMBs have traditionally had high capital ratios.

3.1.2 Case Study of China Merchants Bank

Among small and medium-sized banks, China Merchants Bank (CMB) is an important issuer of wealth management products. In 2012, it accounted for only 3% of total banking assets but 5.2% of WMPs outstanding at year-end and 17.7% of all WMPs issued during the year.[23] The time-variation in CMB's product characteristics will provide further evidence that WMPs are a response to loan-to-deposit rules.

It is useful to note that CMB's loan-to-deposit ratio exhibits much the same patterns as the aggregate JSCB ratios in Figure 6. CMB is one of the twelve joint-stock banks. CMB

[23] Based on data from KPMG, CBRC, and China Merchants Bank.

averaged an end-of-year loan-to-deposit ratio of 74% between 2007 and 2013, just satisfying the 75% cap. When calculated using average balances during the year rather than end-of-year balances, CMB's loan-to-deposit ratio averaged 82% over the same period. The growth in CMB's wealth management products has been dramatic. Annual issuance increased from RMB 0.1 trillion in 2007 to RMB 0.7 trillion in 2008 before reaching almost RMB 5 trillion in 2013. At the end of both 2012 and 2013, CMB had about 83% of its outstanding WMP balances booked off-balance-sheet. Based on notes to the financial statements, figures for earlier years were likely higher.

Count data from Wind indicates that 44% of new WMP batches issued by CMB in 2008 were backed by credit assets and notes. This figure rose to 63% in 2009, consistent with our argument that WMPs were driven by stricter enforcement of loan-to-deposit caps.[24] The use of credit and notes as backing assets has since fallen due to CBRC's rules on (direct) bank-trust cooperation. CMB now backs most of its WMPs with interest rate products, engaging in the counterpart business discussed in Section 2.3.

Further evidence on the importance of loan-to-deposit rules comes from changes in WMP maturity. Figure 8 reveals a sizeable drop in the median maturity of CMB's non-guaranteed products, from just over 4 months in late 2009 to just under 1 month by mid-2011. This drop does not occur for guaranteed WMPs nor is it matched by a drop in the promised annualized yield on non-guaranteed products. Instead, the drop in CMB's non-guaranteed maturity coincides with changes in CBRC's monitoring of loan-to-deposit ratios. Recall from Section 2.2 that CBRC focused on the end-of-year ratio until late 2009, the end-of-quarter ratio until late 2010, and the end-of-month ratio until mid-2011. CMB thus shortened the maturity of its non-guaranteed products as the frequency of CBRC exams increased.

How can maturity be used to thwart more frequent exams? Upon maturity, the principal and interest from a non-guaranteed (off-balance-sheet) WMP are automatically transferred to the buyer's deposit account. A buyer who wants to roll-over his investment then contacts his bank to have the transfer reversed. In the time between the transfer and the reversal though, reserves and deposits rise, lowering the loan-to-deposit ratio observed by CBRC.[25] In the first half of 2011, CMB's non-guaranteed products had a median maturity of just under 1 month which enables the window dressing that thwarts the end-of-month exams. To make

[24] The rise in credit and notes as backing assets between 2008 and 2009 also appears for small and medium-sized banks as a whole (32% to 41%) but not for the Big Four (41% to 37%).

[25] Keeping the automatic deposits as reserves is one approach. Another is to bring loans back on balance sheet in the form of trust beneficiary rights. The data suggest that CMB just recorded reserves between 2009 and 2011.

this point more concrete, we look at the maturity of each non-guaranteed batch relative to its issue date. Approximately 15% of the non-guaranteed batches issued by CMB between January 2008 and December 2010 would have matured near a month-end. This fraction jumped to 40% in early 2011. Arbitraging on maturity became harder in mid-2011 as CBRC began monitoring average daily ratios. Accordingly, Figure 8 shows that CMB's median non-guaranteed maturity has returned to roughly 3 months. The fraction of non-guaranteed batches set to mature near a month-end has also fallen back below 20%.

3.2 Evolution of Total Credit

We have now established that small and medium-sized banks use WMPs to get around stricter loan-to-deposit rules. WMP issuance has grown substantially and, given the high fraction of non-guaranteed WMPs, shadow lending by trust companies has also been able to grow. At the same time, lending by traditional banks has grown too. Commercial banks for which Bankscope has complete data collectively added RMB 40 trillion of new loans between 2007 and 2014, pushing the ratio of traditional lending to GDP from 75% in 2007 to 95% in 2014. Adding this to the growth of the shadow sector estimated in Section 2.4, we get a 35 percentage point increase in the ratio of total credit to GDP. This translates into a roughly 10 percentage point increase in China's credit-to-savings ratio, which we estimate rose from 65% in 2007 to 75% in 2014.

Let us now take a closer look at the Big Four. Section 3.1 showed them to be fairly passive in shadow banking so we are interested to see how, if at all, they have contributed to traditional lending. One indication that the Big Four have become increasingly aggressive in traditional lending comes from Figure 6: their loan-to-deposit ratio was falling prior to 2008 but has been rising ever since. This rise reflects both higher loan growth and lower deposit growth. From 2005 to 2008, loans and deposits at the Big Four grew at annualized rates of 10.9% and 14.1% respectively. From 2008 to 2014, these rates were 16.7% and 12.3% respectively. Why did big banks lend more aggressively against weaker deposit growth exactly when regulators began enforcing loan-to-deposit caps? A common explanation is the two-year RMB 4 trillion stimulus package announced by China's State Council in late 2008. A money multiplier calculation will help us evaluate the adequacy of this explanation. We can then extend the calculation to estimate how much of the overall increase in China's credit-to-savings ratio can be reasonably explained by stimulus alone.

The size of the stimulus is S and the fraction to be financed by the Big Four is q. To finance qS, the Big Four make a one-time transfer of qS from excess reserves to loans. We will treat the Big Four as a closed system, meaning that their lending does not increase deposits at the SMBs. With a currency ratio of c and a reserve ratio of r, the multiplier process increases loans and deposits at the Big Four by $\frac{qS}{1-(1-r)(1-c)}$ and $\frac{(1-c)qS}{1-(1-r)(1-c)}$ respectively. We assume a conservative currency ratio ($c = 0.05$) so as not to understate the effect of the stimulus package on Big Four loans. We then set $r = 0.35$, recalling from Figure 6 that the loan-to-deposit ratio of big banks averaged just over 0.6 between 2005 and 2008. To parameterize q, we note that much of the stimulus package was to be borrowed by local governments. Based on our discussions with CBRC, the central government does not (and did not) pressure the Big Four to finance a disproportionate amount of local government borrowing. The Big Four had a deposit share of roughly 55% in 2007, making it unlikely that they would have been asked to finance more than $q = 0.65$.

The results of our calculation suggest that China's stimulus package can account for up to RMB 6.8 trillion of new loans and up to RMB 6.5 trillion of new deposits at the Big Four since the end of 2008. Taking these effects out, loans and deposits at the Big Four would have grown at annualized rates of 12.9% and 9.8% respectively from 2008 to 2014. The Big Four's loan-to-deposit ratio would have then increased from 0.57 in 2008 to 0.67 in 2014. This counterfactual estimate of what would have happened to the Big Four's loan-to-deposit ratio absent stimulus is similar to what actually happened with the stimulus (an increase from 0.57 in 2008 to 0.70 in 2014 as per Figure 6) so stimulus is an incomplete explanation of why big banks have become less liquid.

Stimulus also does not fully explain the rise in China's aggregate credit-to-savings ratio. To make this statement more precise, suppose the remaining $(1 - q) S$ of stimulus was financed by SMBs. This amount will also go through the multiplier process, except with a lower reserve ratio (call it $\widetilde{r}$) since SMBs have higher loan-to-deposit ratios than the Big Four. We set $\widetilde{r} = 0.15$ based on the average balance data for 2007 in Figure 6. Combining the calculations for the SMBs with the calculations for the Big Four, we find that the stimulus package explains around 40% of the 10 percentage point increase in China's credit-to-savings ratio since 2007.

3.3 Price Manipulation on Interbank Markets

We argue that the Big Four have become strategically less liquid to tighten interbank conditions and put pressure on the shadow banking activities of SMBs. Before formalizing our argument, this section provides evidence that big banks can and do manipulate the interbank market against SMBs. Figure 9 shows an upward trend in monthly average interbank interest rates since 2009 despite fairly large monetary injections by China's central bank. A particularly dramatic spike in interbank rates occurred in the middle of 2013 so we will now dig deeper into this event to see how the Big Four behaved.

Banks in general experienced some liquidity pressure in early June 2013 as companies withdrew deposits to pay taxes and households withdrew ahead of a statutory holiday.[26] Accordingly, the weighted average interbank repo rate rose from 4.64% on June 3 to 9.33% on June 8 before falling back down to 5.37% on June 17.[27] Most of the seasonal pressures seemed to have subsided yet interbank rates rose again on June 20 after the central bank indicated it would not inject extra liquidity. The weighted average repo rate hit 11.57%, with minimum and maximum rates of 4.1% and 30% respectively. For comparison, the minimum and maximum rates on June 3 were 3.87% and 5.32% respectively.

The main net lenders in the interbank repo market on June 20 were China's three policy banks. These banks typically raise money in bond markets to fund economic development projects approved by the central government. The policy banks are almost always net lenders in the interbank repo market but they are usually not the main net lenders. Support for this statement comes from Wind which reports daily net positions by bank type between July 2009 and September 2010. On the 285 (out of 309) trading days where policy banks and big banks were both net lenders, big banks were the main net lender 93% of the time. Moreover, when big banks were the main net lender, their net lending was 4.2 times that of policy banks. In contrast, when policy banks were the main net lender, their net lending was only 1.6 times that of big banks.

The situation was very different on June 20. Big banks were reluctant to lend (Figure 10(a)) and eager to borrow, amassing RMB 50 billion of net borrowing by the end of the trading day. This left policy banks as the main source of interbank liquidity. Figure 10(b) shows a sharp increase in policy bank lending, much of which was absorbed by the big banks. This behavior by the big banks crowded out small and medium-sized banks. For example, as shown in Figure 11, joint-stock banks (JSCBs) paid a lot more for non-policy bank loans on June 20 than they did for policy bank loans. It then stands to reason that JSCBs would have liked a higher share of policy bank lending. Instead, they received 20% of what policy banks lent on June 20, down from an average of 28% over the rest of the month.[28] City and

[26] *The Economist*, "The Shibor Shock," June 22, 2013.

[27] We focus on the interbank repo market rather than the uncollateralized money market since the former is much bigger than the latter.

[28] For completeness, the overnight and 7 day maturities shown in Figure 11 were almost 94% of JSCB borrowing on June 20. They were also 100% of JSCB borrowing from policy banks on this date. There were no major differences in the haircuts imposed by policy banks versus other lenders.

rural banks also faced large price differentials between policy and non-policy bank loans. However, their share of policy bank lending on June 20 was 22%, well below an average of 47% over the rest of the month.

Were big banks borrowing on June 20 because they really needed liquidity? Two pieces of evidence suggest no. First, their ratio of repo lending to repo borrowing was 0.7, with 71% of the loans *not* directed towards policy or other big banks. If the Big Four were in dire need of liquidity on June 20, we would expect to see very little outflow. Second, the repo activities of big banks involved a maturity mismatch. Excluding transactions within the Big Four, overnight trades accounted for 96% of big bank borrowing but only 83% of big bank lending to non-policy banks. Roughly 80% of policy bank lending to banks outside the Big Four was also at the overnight maturity. If big banks really needed liquidity on June 20, we would expect the maturity of their lending to be closer to the maturity of their borrowing. Instead, it was closer to the maturity offered by policy banks to borrower groups that policy banks and big banks had in common.

Figure 12 shows that big banks also commanded an abnormally high interest rate spread on June 20. In particular, their weighted average lending rate was 266 basis points above their weighted average borrowing rate. This is high relative to other banks: city banks and JSCBs commanded spreads of 46 and 113 basis points respectively. It is also high relative to other days in the sample: on any other day in June 2013, the spread commanded by big banks was between -40 and 58 basis points.

Finally, we look at dispersion in the lending rates charged by the Big Four and find evidence of collusive pricing.[29] In June 2013, the average daily coefficient of variation for overnight lending rates offered by big banks was 62% of the average coefficient for JSCBs and 29% of the average coefficient for city banks. These figures were 61% and 21% respectively on June 20. The data thus reveals more uniform pricing among big banks than among SMBs.

A common narrative is that China's interbank market tightened on June 20 because the central bank wanted to shock and discipline it. Our evidence challenges this narrative in two ways. First, the policy banks were lending a lot of money at fairly low interest rates. Given their political nature, they would not have behaved this way had the central bank really wanted to shock the market. Second, the Big Four were manipulating the interbank market by absorbing liquidity and intermediating it to SMBs at much higher interest rates. A potentially useful way to summarize the importance of the Big Four is to report how

[29] We exclude lending rates charged to policy banks given the proximity of policy banks to the government.

the overnight repo rate (ONR) correlates with the amount lent by big banks versus policy banks. In June 2013, the correlation between the ONR and the amount lent by big banks to JSCBs was -0.38. In contrast, the correlation between the ONR and the amount lent by policy banks to JSCBs was 0.13. A similar pattern emerges for lending to other SMBs: the correlation between the ONR and the amount lent by big banks to other SMBs was -0.62 while the correlation between the ONR and the amount lent by policy banks to other SMBs was only -0.22. The much stronger (negative) correlation between interbank lending by big banks and the interbank rate repo rate captures that big banks do indeed have the ability to manipulate interbank prices.

4 A Unifying Theory of the Facts

The previous section established a set of facts about China's banking system. We argued that stricter loan-to-deposit rules triggered shadow banking among small and medium-sized banks. We also argued that big banks have become less liquid and are manipulating the interbank market. The end result has been an increase in total credit and an increase in interbank interest rates. Stricter regulation has thus been entirely counter-productive.

We now build a banking model that connects all the facts. In particular, we connect interbank manipulation by the Big Four with shadow banking by the SMBs and show that the net effect of a tighter loan-to-deposit cap is indeed more credit and a higher interbank rate. Our model has three main ingredients: (i) liquidity shocks, (ii) an interbank market for reserves, and (iii) heterogeneity in interbank market power. The third ingredient is motivated by the evidence in Section 3.3, namely the fact that the Big Four can change interbank prices if they so choose. To isolate the contribution of this ingredient and expound the mechanisms behind our quantitative results in Section 5, we proceed in steps. Section 4.1 begins by describing an environment without heterogeneity. Section 4.2 then shows that this environment only delivers some of the facts, namely the rise of shadow banking after stricter loan-to-deposit rules but not the increase in total credit or the increase in interbank interest rates. Heterogeneity in interbank market power is introduced in Section 4.3 and shown to deliver a much more comprehensive picture in Section 4.4.

4.1 Benchmark Model

There are three periods, $t \in \{0, 1, 2\}$, and a unit mass of risk neutral banks, $j \in [0, 1]$. The economy is endowed with $X > 0$ units of funding. Let X_j denote the amount of funding obtained by bank j at $t = 0$, where $\int X_j dj = X$.

Each bank can invest in a project which returns $(1 + i_A)^2$ per unit invested. Projects are long-term, meaning that they run from $t = 0$ to $t = 2$ without the possibility of liquidation at $t = 1$. However, banks are subject to short-term idiosyncratic liquidity shocks which must be paid off at $t = 1$. More precisely, bank j must pay $\theta_j X_j$ at $t = 1$ in order to continue operation. The exact value of θ_j is drawn from a two-point distribution:

$$\theta_j = \begin{cases} \theta_\ell & prob.\ \pi \\ \theta_h & prob.\ 1 - \pi \end{cases}$$

where $0 < \theta_\ell < \theta_h < 1$ and $\pi \in (0, 1)$. Each bank learns the realization of its shock in $t = 1$. Prior to that, only the distribution is known.

To make the model more concrete, we can interpret the liquidity shocks through the lens of Diamond and Dybvig (1983). In particular, banks attract funding at $t = 0$ by offering liquidity services (e.g., deposits and/or WMPs) to owners of the funding endowment (e.g., households). The shock θ_j then represents the fraction of households that withdraw their deposits and WMPs from bank j at $t = 1$.

Let us now spell out the liquidity services provided by banks. A dollar deposited at $t = 0$ becomes $1 + i_B$ if withdrawn at $t = 1$ and $(1 + i_D)^2$ if withdrawn at $t = 2$. A WMP involves the same returns plus an additional return ξ_j. To ease the exposition, suppose ξ_j only accrues if the WMP is held until $t = 2$. In Diamond and Dybvig (1983), banks choose deposit rates to achieve optimal risk-sharing for households. In China, deposit rates are essentially set by the government through binding ceilings (Section 2.2). We therefore treat i_B and i_D as exogenous.[30] Without loss of generality for the analytical results, we can then normalize $i_B = i_D = 0$ and interpret the other interest rates in the model as spreads relative to this normalization.

Let D_j denote the funding attracted by bank j in the form of traditional deposits. The funding attracted in the form of WMPs is W_j, with $X_j \equiv D_j + W_j$. In the data, deposits and WMPs co-exist despite the fact that WMPs pay higher returns (even after potential risk adjustments). This suggests that deposits have a convenience value which stops households

[30] See Diamond and Kashyap (2015) for another Diamond-Dybvig environment with exogenously given deposit rates. They motivate by saying that households have an exogenous outside option which puts a floor on deposit rates and, without competition between banks, the floor binds. In our model, the government just sets the ceiling equal to the outside option.

from switching entirely to WMPs. We will thus model D_j and W_j as continuous functions. Consider for example:

$$W_j = \omega \xi_j \tag{1}$$

$$D_j = X - \rho_1 \xi_j - \rho_2 \overline{\xi} \tag{2}$$

where $\overline{\xi}$ denotes the average WMP return and ω, ρ_1, and ρ_2 are non-negative constants. Each individual bank takes $\overline{\xi}$ as given. A symmetric equilibrium requires $\xi_j = \overline{\xi}$ and $X_j = X$ for all j so we must impose:

$$\omega = \rho_1 + \rho_2$$

This then allows us to write bank j's funding share as:

$$X_j = X + \rho_2 \left(\xi_j - \overline{\xi}\right) \tag{3}$$

The key feature of (1) and (2) is that higher WMP returns prompt a partial substitution from deposits to WMPs. A simple microfoundation for the specific functional forms used is sketched in Appendix B.

We now describe how banks use their funds. The maturity mismatch between investment projects and liquidity shocks introduces a role for reserves (i.e., precautionary savings which can be used to pay realized liquidity shocks). The division of X_j into investment and reserves is chosen at $t = 0$. Let $R_j \in [0, X_j]$ denote the reserve holdings of bank j. If $\theta_j < \frac{R_j}{X_j}$, then bank j has a reserve surplus at $t = 1$. Otherwise ($\theta_j > \frac{R_j}{X_j}$), it has a reserve shortage.

An interbank market exists at $t = 1$ to redistribute reserves across banks. The interest rate in this market is i_L. Banks are atomistic so they take i_L as given when making decisions. However, i_L adjusts to clear the interbank market. Interbank lenders (borrowers) are banks with reserve surpluses (shortages) at $t = 1$. Some lending may also be done by the central bank so we introduce a supply of external funds $\Psi\left(i_L\right) \equiv \psi i_L$, where $\psi > 0$. We focus on symmetric equilibrium, in which case R_j and X_j are the same across banks. The condition for interbank market clearing is then:

$$R_j + \psi i_L = \overline{\theta} X \tag{4}$$

where $\overline{\theta} \equiv \pi \theta_\ell + (1 - \pi)\, \theta_h$ is the average liquidity shock. Total credit in this economy is the total amount of funding invested in projects (i.e., $1 - R_j$).

We now allow for the possibility of a government-imposed loan limit on each bank. This limit can also be viewed as a liquidity rule which says the ratio of reserves to on-balance-sheet funding must be at least $\alpha \in (0,1)$. Given the structure of our model, reserves are meant to be used at $t=1$ so enforcement of the liquidity rule is confined to $t=0$. If the government does not enforce a liquidity rule, then $\alpha = 0$.

Whereas deposits must be booked on-balance-sheet, banks can choose where to manage WMPs and the projects financed by those WMPs. If fraction $\tau_j \in [0,1]$ is managed in an off-balance-sheet vehicle, then bank j's reserve holdings only need to satisfy:

$$R_j \geq \alpha \left(X_j - \tau_j W_j\right) \tag{5}$$

The use of off-balance-sheet vehicles constitutes regulatory arbitrage.[31] We can now employ our model to study whether regulatory arbitrage is an equilibrium response to changes in liquidity rules. To make the policy experiment concrete, suppose the government moves from $\alpha = 0$ to $\alpha = \overline{\theta}$. Our goal here is to build intuition; Section 5 will fit the starting and ending values of α to data.

4.2 Results for Benchmark Model

The expected profit of bank j at $t=0$ is:

$$\Upsilon_j \equiv (1+i_A)^2 (X_j - R_j) + (1+i_L) R_j - \left[X_j + i_L \overline{\theta} X_j + \left(1-\overline{\theta}\right) \xi_j W_j\right] - \tfrac{\phi}{2} X_j^2 \tag{6}$$

where W_j and X_j are given by (1) and (3) respectively. The first term in (6) is revenue from investment. The second term is potential revenue from reserves. The third term is the bank's expected funding cost. The fourth term is a quadratic operating cost (with $\phi > 0$) which will play a minimal role until Section 4.3.

The representative bank chooses the attractiveness of its WMPs ξ_j, the intensity of its off-balance-sheet activities $\tau_j \in [0,1]$, and its reserve holdings R_j to maximize Υ_j subject to the liquidity rule in (5). The Lagrange multiplier on (5) is the shadow cost of reserves. We denote it by μ_j. The multipliers on $\tau_j \geq 0$ and $\tau_j \leq 1$ are denoted by η_j^0 and η_j^1 respectively. The first order conditions with respect to R_j, τ_j, and ξ_j are then:

$$\mu_j = (1+i_A)^2 - (1+i_L) \tag{7}$$

$$\eta_j^1 = \eta_j^0 + \alpha \mu_j W_j \tag{8}$$

[31] See, for example, Adrian et al (2013) who define regulatory arbitrage as "a change in structure of activity which does not change the risk profile of that activity, but increases the net cash flows to the sponsor by reducing the costs of regulation."

$$\xi_j = \frac{\left(1-\overline{\theta}\right) i_L + (1-\alpha)\,\mu_j - \phi X_j}{2\left(1-\overline{\theta}\right)} \times \frac{\rho_2}{\omega} + \frac{\alpha \mu_j \tau_j}{2\left(1-\overline{\theta}\right)} \tag{9}$$

The first term in equation (9) captures what we will call the competitive motive for WMP issuance. If this term is positive, then bank j wants to offer higher WMP returns in order to attract more funding. Recall that bank j's total funding, X_j, is given by equation (3). Each bank takes $\overline{\xi}$ as given so increasing ξ_j relative to $\overline{\xi}$ increases X_j. The second term in equation (9) captures what we will call the regulatory arbitrage motive for WMP issuance. In the absence of a liquidity rule ($\alpha = 0$), there is no regulatory arbitrage motive. There is also no such motive when the interbank rate is high enough to make the shadow cost of reserves (μ_j) zero.[32]

Proposition 1 *Suppose $\phi < \overline{\phi}$ where $\overline{\phi}$ is a positive threshold. If $\alpha = 0$, then $\xi_j = 0$ if and only if $\rho_2 = 0$.*

Proposition 1 establishes that there is always a competitive motive for WMP issuance when $\rho_2 > 0$ and operation costs are modest ($\phi < \overline{\phi}$). This is because each bank perceives its WMPs as cutting at least partly into the funding share of other banks. If $\rho_2 = 0$, then each bank perceives its WMPs as cutting one-for-one into its own funding share. Therefore, with $\rho_2 = 0$, a regulatory arbitrage motive is both necessary and sufficient for WMP issuance. The data show that very few WMPs were issued before CBRC's enforcement of loan-to-deposit caps, making the appropriate starting point either $\rho_2 = 0$ or the combination of $\rho_2 > 0$ and $\phi \geq \overline{\phi}$. We start with $\rho_2 = 0$ since it is analytically more convenient but, as will be seen later, $\rho_2 > 0$ with ϕ sufficiently high delivers the same intuition.[33]

Proposition 2 *Suppose $\rho_2 = 0$. There is a unique $\overline{\alpha} \in \left[0, \overline{\theta}\right)$ such that $\xi_j = 0$ if $\alpha \leq \overline{\alpha}$ and $\xi_j > 0$ with $\tau_j = 1$ otherwise.*

In words, Proposition 2 says that sufficiently strict liquidity regulation (i.e., increasing α above $\overline{\alpha}$) triggers the issuance of off-balance-sheet WMPs. The benchmark model can therefore account for the rise of shadow banking. The incentive to issue WMPs does not come from competition: with $\rho_2 = 0$, the bank is simply substituting within its own liabilities. Instead, WMPs are issued because they can be booked off-balance-sheet, away from the binding liquidity rule.

[32] The competitive motive can also be interpreted as a type of regulatory arbitrage, where the regulation being circumvented is the uncompetitive ceiling on deposit rates. Using the term "regulatory arbitrage" with reference to both liquidity rules and deposit rate rules is confusing. Since Section 3.1 showed that liquidity rules were the trigger for China's shadow banking, we reserve the term for liquidity rules.

[33] See Proposition 3 and Section 5.

Proposition 3 *For any $\rho_2 \geq 0$, the interbank rate in the benchmark model is highest at $\alpha = 0$. Moving from $\alpha = 0$ and $\rho_2 = 0$ to $\alpha > 0$ and $\rho_2 > 0$ will also not generate an increase in the interbank rate or an increase in total credit.*

While the benchmark model is useful for understanding the motives behind WMP issuance, Proposition 3 shows that introducing a liquidity rule into this model will always lead to a decrease in the interbank rate. Total credit $(1 - R_j)$ must then also fall given (4).[34] Proposition 3 is basically the market mechanism at work. Suppose there is no government intervention ($\alpha = 0$). At low interbank rates, price-taking banks will rely on the interbank market for liquidity instead of holding their own reserves. In a Walrasian market, all banks are price-takers so there will be liquidity demand at $t = 1$ but no liquidity supply. This cannot be an equilibrium. Therefore, the interbank rate must be high to substitute for government intervention.

4.3 Adding Heterogeneity in Market Power

We now extend the benchmark model to include a big bank. By definition of being big, the big bank is *not* an interbank price-taker.

We keep the continuum of small banks, $j \in [0, 1]$, and index the big bank by k. WMP demands are $W_j = \omega\xi_j$ and $W_k = \omega\xi_k$, similar to equation (1). The funding attracted by each bank is an augmented version of equation (3), namely:

$$X_j = 1 - \delta_0 + \delta_1 \left(\xi_j - \xi_k\right) + \delta_2 \left(\xi_j - \overline{\xi_j}\right) \tag{10}$$

$$X_k = \delta_0 + \delta_1 \left(\xi_k - \overline{\xi_j}\right) \tag{11}$$

where $\overline{\xi_j}$ is the average return on small bank WMPs and total funding in the economy has been normalized to $X = 1$. Small banks take $\overline{\xi_j}$ and ξ_k as given, along with being interbank price-takers. In a symmetric equilibrium, $\xi_j = \overline{\xi_j}$. The big bank does *not* take $\overline{\xi_j}$ as given.

Since the big bank is effectively an interbank price-setter, the interbank rate will depend on the big bank's realized liquidity shock. This makes the big bank's shock an aggregate shock so Appendix C shows that adding aggregate shocks to the benchmark model with only small banks does not change Proposition 3.

Let i_L^s denote the interbank rate when the big bank realizes θ_s, where $s \in \{\ell, h\}$. To make our main points, it will be enough for the big bank to affect the expected interbank

[34] If $\psi = 0$, then total credit will be constant. Either way, there cannot be a credit boom.

rate, $i_L^e \equiv \pi i_L^\ell + (1-\pi)\, i_L^h$. We can therefore simplify the exposition by fixing $i_L^\ell = 0$ and letting i_L^e move with i_L^h.[35] In Sections 4.1 and 4.2, we used market clearing to pin down the endogenous interbank rate. We will use a similar approach here. In particular, when the big bank gets a high liquidity shock, the condition for interbank market clearing is:

$$R_j + R_k + \psi i_L^h = \overline{\theta} X_j + \theta_h X_k \tag{12}$$

The left-hand side captures the supply of liquidity while the right-hand side captures the demand for liquidity.

At $t = 0$, the big bank's expected profit is:

$$\Upsilon_k \equiv (1+i_A)^2 (X_k - R_k) + \left[1 + (1-\pi)\, i_L^h\right] R_k - \left[X_k + (1-\pi)\, i_L^h \theta_h X_k + \left(1-\overline{\theta}\right) \omega \xi_k^2\right] - \frac{\phi}{2} X_k^2$$

The interpretation is similar to equation (6): the first term is revenue from investment, the second term is the expected potential revenue from reserves, the third term is the big bank's expected funding cost, and the fourth term is an operating cost.[36]

The big bank chooses R_k, τ_k, and ξ_k to maximize Υ_k subject to three sets of constraints. First are the aggregate constraints, namely funding shares as per (10) and (11) and market clearing as per (12). Note that the market clearing equation connects R_k and i_L^h. Therefore, we can think of the big bank as choosing i_L^h with R_k determined by (12), rather than the other way around.

The second set of constraints comes from the first order conditions of small banks. The representative small bank solves essentially the same problem as before. Its objective function is still given by (6) but with $(1-\pi)\, i_L^h$ as the interbank rate and X_j as per equation (10). In the benchmark model, we used $\rho_2 = 0$ in equation (10) to capture the empirical fact that small banks issued virtually no WMPs at low α. The counterpart here is $\delta_1 + \delta_2 = 0$ in equation (10), making the small bank first order conditions:

$$\mu_j \left[R_j - \alpha \left(X_j - \omega \xi_j\right)\right] = 0 \text{ with complementary slackness} \tag{13}$$

$$\mu_j = (1+i_A)^2 - \left[1 + (1-\pi)\, i_L^h\right] \tag{14}$$

[35] The proofs will verify that $i_L^\ell = 0$ does not result in a liquidity shortage when the big bank realizes θ_ℓ.

[36] We assume the same "risk-adjusted" return to investing in state sector versus private sector firms. The private sector is more productive than the state sector but, at least politically, lending to the private sector is riskier. Some anecdotal evidence can be found in Dobson and Kashyap (2006).

$$\xi_j = \frac{\alpha \mu_j}{2\left(1-\overline{\theta}\right)} \tag{15}$$

Notice that equation (15) captures essentially the same regulatory arbitrage motive as equation (9). It will be useful to deviate from the benchmark model in small steps, hence the use of $\delta_1 + \delta_2 = 0$ rather than $\delta_1 + \delta_2 > 0$. However, we will allow for $\delta_1 + \delta_2 > 0$ (with appropriate ϕ) in the calibration and show that the qualitative results are unchanged.

The last set of constraints on the big bank's problem are inequality constraints, namely the liquidity rule and non-negativity conditions:

$$R_k \geq \alpha \left(X_k - \tau_k W_k\right)$$

$$\tau_k \in [0, 1]$$

$$\xi_k \geq 0$$

$$\mu_j \geq 0$$

Each inequality constraint can be in one of two cases: binding or slack. In the data, big banks are not constrained by the liquidity rule. They are also less involved in off-balance-sheet activities than small banks. To capture this, we will look for a solution with $R_k > \alpha X_k$ and $\tau_k = 0$.[37] We will also require $\xi_k = 0$ at $\alpha = 0$ since very few WMPs were issued before CBRC's enforcement of loan-to-deposit caps. Finally, we will work with $\mu_j > 0$ to capture the fact that small banks are constrained by liquidity rules.

When the big bank is unconstrained by the liquidity rule, it has no regulatory arbitrage motive for WMP issuance. Therefore, getting $\xi_k = 0$ at $\alpha = 0$ just requires disciplining any competitive motive the big bank may have. As was the case for small banks in Proposition 1, there are two ways to do this. The first is to shut down the competitive motive altogether (i.e., $\delta_1 = 0$). The second is to allow for a sufficiently high operating cost (i.e., $\delta_1 > 0$ but with ϕ sufficiently positive). We will consider both cases to better understand the role that big banks can play in this economy. The first case, $\delta_1 = 0$, implies a fixed funding share for the big bank (i.e., $X_k = \delta_0$). The second case, $\delta_1 > 0$, makes the big bank's funding share endogenous. For the second case, we will set ϕ so that the equilibrium ξ_k is exactly zero at $\alpha = 0$ (as opposed to ξ_k being constrained by zero). We will be interested to see how, if at all, an endogenous funding share affects the big bank's decision-making.

[37] The big bank is technically indifferent between any $\tau_k \in [0, 1]$ if its rule is slack. We consider $\tau_k = 0$ for analytical convenience and because finding an equilibrium with $\tau_j = 1$ and $\tau_k = 0$ will be enough to capture the empirical fact that SMBs have a higher intensity of non-guaranteed activity.

4.4 Results for Extended Model

An equilibrium is characterized by the first order conditions from the small bank problem, the first order conditions from the big bank problem, and interbank market clearing.

Proposition 4 *Suppose $\alpha = 0$. Set $\delta_1 + \delta_2 = 0$ to get $\xi_j = 0$. Also set either $\delta_1 = 0$ or $\delta_1 > 0$ with ϕ sufficiently positive to get $\xi_k = 0$. If i_A lies within an intermediate range, the equilibrium at $\alpha = 0$ involves $\mu_j > 0$, $R_j = 0$, and $R_k > 0$.*

As in the benchmark model, a liquidity rule is not needed for the banking system to be liquid. In contrast to the benchmark though, liquidity is now held disproportionately by the big bank: small banks invest all their funding in projects and rely on the interbank market to honor short-term obligations. The big bank's willingness to hold liquidity reflects its status as an interbank price-setter. In particular, the big bank understands that not holding enough liquidity will increase its funding costs should it experience a high liquidity shock.

We now conduct the same policy experiment as we did in the benchmark model: the government increases the liquidity rule from $\alpha = 0$ to $\alpha = \overline{\theta}$. As discussed in Section 2, the government's objective could be to stifle small banks by forcing them to hold some of their own liquidity rather than expecting a subsidy from big bank liquidity holdings. Proposition 4 established $\mu_j > 0$ at $\alpha = 0$ so forcing small banks to shift from investment projects to reserves is indeed a tax on them. As shown next, the introduction of a big bank overcomes the shortcomings of the benchmark model and accounts for the rise in interbank rates:

Proposition 5 *Keep $\delta_1 + \delta_2 = 0$ as in Proposition 4. The following are sufficient for $\alpha = \overline{\theta}$ to generate a higher interbank rate than $\alpha = 0$ while preserving slackness of the big bank's liquidity rule, bindingness of the small bank liquidity rule ($\mu_j > 0$), and feasibility of $i_L^\ell = 0$:*

1. *Suppose $\delta_1 = 0$ so that the big bank's funding share is fixed. The sufficient conditions are: π sufficiently high, θ_ℓ and $\frac{\psi}{\omega}$ sufficiently low, and i_A within an intermediate range.*

2. *Suppose $\delta_1 = \omega > 0$ so that the big bank's funding share is endogenous. Also set ϕ so that ξ_k is exactly zero at $\alpha = 0$. The sufficient conditions are: π sufficiently high, θ_ℓ and $\frac{\psi}{\omega}$ sufficiently low, and i_A and δ_0 within intermediate ranges.*

There is a non-empty set of parameters satisfying the sufficient conditions in both 1 and 2. All else constant, the model with an endogenous funding share generates a larger increase in the interbank rate than the model with a fixed funding share.

To explain the content of Proposition 5, it will be useful to summarize all the forces behind the big bank's choice of i_L^h. From the big bank's objective function:

$$\frac{\partial \Upsilon_k}{\partial i_L^h} \propto \underbrace{R_k - \theta_h X_k}_{\text{direct motive}} - \underbrace{\left[\frac{(1+i_A)^2 - 1}{1-\pi} - i_L^h\right]\frac{\partial R_k}{\partial i_L^h}}_{\text{reallocation motive}} + \underbrace{\left[\frac{(1+i_A)^2 - 1 - \phi X_k}{1-\pi} - \theta_h i_L^h\right]\frac{\partial X_k}{\partial i_L^h}}_{\text{funding share motive}} \tag{16}$$

The equilibrium i_L^h solves $\frac{\partial \Upsilon_k}{\partial i_L^h} = 0$. We will first explain the three motives identified in (16). We will then explain how these motives vary with α in order to understand why moving from $\alpha = 0$ to $\alpha = \overline{\theta}$ generates a higher interbank rate.

The first motive is what we call the direct motive. The big bank has reserves R_k and a funding share X_k. Its net reserve position when hit by a high liquidity shock is therefore $R_k - \theta_h X_k$. Each unit of reserves is valued at an interest rate of i_L^h when the big bank's shock is high so, on the margin, an increase in i_L^h changes the big bank's profits by $R_k - \theta_h X_k$. The second motive is what we call the reallocation motive. The idea is that changes in i_L^h also affect how many reserves the big bank needs to hold in a market clearing equilibrium. If $\frac{\partial R_k}{\partial i_L^h} < 0$, then an increase i_L^h elicits enough liquidity from other sources to let the big bank reallocate funding from reserves to investment. On the margin, the value of this reallocation is the shadow cost of reserves, hence the coefficient on $\frac{\partial R_k}{\partial i_L^h}$ in (16). The third motive is what we call the funding share motive. The idea is that changes in i_L^h also affect how much funding the big bank attracts when funding shares are endogenous. If $\frac{\partial X_k}{\partial i_L^h} > 0$, then an increase in i_L^h curtails the WMP activities of small banks by enough to boost the big bank's funding share. The coefficient on $\frac{\partial X_k}{\partial i_L^h}$ in (16) captures the marginal value of a higher funding share for the big bank. We will discuss this coefficient in more detail below.

To gain some insight into how changes in α will affect the solution to $\frac{\partial \Upsilon_k}{\partial i_L^h} = 0$ through each motive, let's start with the case of fixed funding shares ($\delta_1 = 0$). From market clearing:

$$R_k - \theta_h X_k \overset{\delta_1 = 0}{=} \overline{\theta}(1-\delta_0) - \psi i_L^h - \underbrace{\alpha\left(1 - \delta_0 - \frac{\alpha\omega(1-\pi)}{2\left(1-\overline{\theta}\right)}\left[\frac{(1+i_A)^2 - 1}{1-\pi} - i_L^h\right]\right)}_{R_j \text{ as per small bank FOCs in (13) to (15)}} \tag{17}$$

For a given value of i_L^h, the magnitude of the direct motive in (17) depends on α through the reserve holdings of small banks. There are two competing effects. On one hand, higher α forces small banks to hold more reserves per unit of on-balance-sheet funding. On the other hand, higher α compels small banks to engage in regulatory arbitrage (via ξ_j) and move funding off-balance-sheet. The net effect is ambiguous so we must look beyond the direct motive to explain Proposition 5.

With fixed funding shares, the only other motive is the reallocation motive:

$$\left.\frac{\partial R_k}{\partial i_L^h}\right|_{\delta_1=0} = -\psi - \frac{\alpha^2\omega\,(1-\pi)}{2\left(1-\overline{\theta}\right)} < 0 \tag{18}$$

This expression is negative for two reasons. First, a higher interbank rate will attract more external liquidity, allowing the big bank to hold fewer reserves. This is captured by the first term in (18). Second, small banks will increase their reserves when the interbank rate increases, also allowing the big bank to hold fewer reserves. This is captured by the second term in (18). The effect of i_L^h on R_j works through the regulatory arbitrage motive of small banks: there is less incentive to circumvent a liquidity regulation when the price of liquidity is expected to be high. We can also see that the effect of i_L^h on R_j strengthens with α. This is both because R_j is more responsive to changes in ξ_j at high α (see equation (13)) and because ξ_j is more responsive to changes in i_L^h at high α (see equations (14) and (15)). These results help us understand the first bullet in Proposition 5: when funding shares are fixed, high α makes it easier for the big bank to use high interbank rates to make small banks share the burden of keeping the system liquid.

Does the same intuition extend to the case of endogenous funding shares? No because:

$$\left.\frac{\partial R_k}{\partial i_L^h}\right|_{\delta_1=\omega} = -\psi + \frac{\alpha\omega\pi\,(\theta_h-\theta_\ell)\,(1-\pi)}{2\left(1-\overline{\theta}\right)} \tag{19}$$

Higher i_L^h still decreases ξ_j but now the decrease in ξ_j decreases how much funding small banks attract (X_j) and therefore how many reserves they need to hold. This effect is strong enough to make the second term in (19) positive, in contrast to (18). We must therefore turn to the funding share motive to explain the second bullet in Proposition 5. Note:

$$\left.\frac{\partial X_k}{\partial i_L^h}\right|_{\delta_1=\omega} = \frac{\alpha\omega\,(1-\pi)}{2\left(1-\overline{\theta}\right)} > 0 \tag{20}$$

We already know that an increase in i_L^h decreases ξ_j which then decreases X_j. Total funding is normalized to one so the decrease in X_j implies an increase in X_k, culminating in (20) being positive. We also know that ξ_j is more responsive to i_L^h at high α so the magnitude of (20) increases with α. It is therefore easier for the big bank to increase its funding share by increasing i_L^h when α is high. To complete the intuition, let us reconcile the big bank's desire to increase its funding share when α is high with the existence of convex operating costs. Return to the coefficient on $\frac{\partial X_k}{\partial i_L^h}$ in (16). All else constant, moving from $\alpha = 0$ to $\alpha = \overline{\theta}$ will trigger regulatory arbitrage by small banks. The resulting increase in ξ_j will erode the big bank's funding share, thereby decreasing the marginal operating cost ϕX_k.

We can now understand the second bullet in Proposition 5 as follows: when funding shares are endogenous, high α makes it easier for the big bank to use high interbank rates to stop small banks from encroaching on its funding share. The last part of Proposition 5 establishes that sizeable increases in the interbank rate are most consistent with this sort of asymmetric competition, wherein the big bank uses its interbank market power to fend off competition from small banks and their off-balance-sheet activities. The next proposition shows that our model also delivers the other empirical facts:

Proposition 6 *Invoke the parameter conditions from Proposition 5 and define the loan-to-deposit ratio as the ratio of investment to on-balance-sheet funding. Total credit increases and the loan-to-deposit ratios of big and small banks converge when we move from $\alpha = 0$ to $\alpha = \overline{\theta}$. Moreover, $\xi_j > \xi_k$ at $\alpha = \overline{\theta}$ with $\xi_k > 0$ if and only if funding share is endogenous. This is in contrast to $\xi_j = \xi_k = 0$ at $\alpha = 0$.*

It may now be useful to recap the intuition for our results, highlighting along the way the channels through which total credit increases. Small banks move into off-balance-sheet WMPs after liquidity rules tighten. Once there, they can also offer interest rates well above the rates permitted for traditional deposits. All else constant, this poaches funding from the big bank. Recall that the big bank internalizes the effect of reserve holdings on the interbank market. Therefore, compared to small banks, it invests less at $t = 0$ per unit of funding attracted. The reallocation of funding from deposits at the big bank to high-return WMPs at the small banks thus increases total credit. This is one of two channels.

The second channel stems from how the big bank responds to its loss of funding. One way for the big bank to respond is by offering its own WMPs with high interest rates. Naturally, this is costly because of the high rates. Another way for the big bank to respond is to use the interbank market. Small banks have less incentive to skirt liquidity rules if they expect the price of liquidity to be high. All else constant, the interbank market at $t = 1$ will be less liquid and the expected interbank rate will rise if the big bank holds fewer reserves at $t = 0$. The big bank can thus manipulate the interbank market to make small banks scale back their issuance of WMPs. While this strategy by the big bank curbs *some* of the initial increase in total credit, it also boosts credit directly because the big bank shifts from reserves to investment at $t = 0$. Notice that the big bank's strategy also contributes directly to the rise in its loan-to-deposit ratio.

5 Calibration Results

This section presents a calibration to show that the forces in our model are quantitatively important. The starting point is 2007, just prior to China's adoption of stricter liquidity rules. The ending point is 2014 which is the most recent year of complete data.

We take the time from $t = 0$ to $t = 2$ to be a quarter. The average benchmark interest rate for three-month deposits in China is 2.6% annualized so we set $(1 + i_D)^2 = 1.026$. The benchmark demand deposit rate averages around 0.4% so we set $(1 + i_B)^2 = 1.004$. The average benchmark interest rate for loans with a maturity of less than six months is 5.6%. Banks can offer a discount of up to 10% on the benchmark loan rate so we set $(1 + i_A)^2 = 1.05$.

The initial liquidity rule is set to $\alpha = 0.14$ to match the observed loan-to-deposit ratio of JSCBs in 2007. We use the ratio based on average balances rather than end-of-year balances, as per Figure 6. The policy experiment is then an increase from $\alpha = 0.14$ to $\alpha = 0.25$, capturing CBRC's stricter enforcement of the 75% loan-to-deposit cap and the complementary increase in reserve requirements.

The Big Four accounted for around 55% of all deposits in China in 2007 so we set $\delta_0 = 0.55$. We then calibrate the competition parameters (δ_1 and δ_2) and the WMP demand parameter (ω) to match funding outcomes in 2014. We saw in Section 2.4 that WMPs were around 25% of GDP at the end of 2014. This is equivalent to 15% of total savings. Small and medium-sized banks accounted for roughly two-thirds of WMPs so we will target $W_j = 0.10$ and $W_k = 0.05$ for 2014. We will also target a funding share of $X_k = 0.45$ for 2014 since the Big Four accounted for roughly 45% of all savings (i.e., traditional deposits plus WMPs) in that year.

We allow big and small banks to have different operating cost parameters, ϕ_k and ϕ_j. China has around 200 commercial banks so a big bank is on average 40 times as large as a small bank (i.e., $\frac{0.45}{4}/\frac{0.55}{196}$). We set $\phi_j = 40\phi_k$ so that marginal operating costs are the same across banks.[38] We then calibrate ϕ_k to match a loan-to-deposit ratio of 0.70 for the Big

[38] In robustness checks, we found that cutting the $\frac{\phi_j}{\phi_k}$ ratio by half to exclude some of the smallest banks and re-calibrating the model generates very similar results.

Four in 2014. We will check that the resulting operating cost parameters are high enough to deliver negligible WMP issuance in 2007.

For the low liquidity shock, the analytical results in Proposition 5 point to a small magnitude (θ_ℓ) and a high probability (π) so we use $\theta_\ell = 0$ and $\pi = 0.75$. Proposition 5 also points to a low external liquidity parameter (ψ) so we use $\psi = 0.5$. We will allow $i_L^\ell = i_B > 0$ in the calibration since surplus reserves can earn a small interest rate from the central bank. We then redefine $\Psi(i_L) \equiv \psi(i_L - i_B)$ to preserve $\Psi\left(i_L^\ell\right) = 0$. Lastly, we calibrate the average liquidity shock ($\overline{\theta} \equiv \pi\theta_\ell + (1-\pi)\theta_h$) to get an average interbank rate of 3.6% when $\alpha = 0.25$. The 3.6% target is the weighted average seven-day interbank repo rate in 2014. The seven-day rate is the longest maturity for which there is significant trading volume. It is difficult to target shorter-term (e.g., overnight) repo rates since we are working with a two-period model and each period must be long enough to match reasonable data on loan returns (i_A). This is just a level effect though: the correlation between the overnight and seven-day repo rates is around 0.95.

The results are summarized in Table 2.[39] Our model generates most of the rise in WMPs between 2007 and 2014. It also generates half of the increase in the average seven-day interbank repo rate. Since yearly averages can mask some of the most severe events, it is also useful to consider the peak interbank rates observed before and after CBRC's enforcement action (10.1% and 11.6% respectively, as measured by daily averages). Of this 150 basis point increase in peak rates, our model delivers 90 basis points. Returning to Table 2, the model also delivers more than half of the decrease in the Big Four's funding share. We also obtain a large increase in the Big Four's loan-to-deposit ratio, although the increase is somewhat larger than what we observe in the data. Finally, we obtain a sizeable 3.2 percentage point increase in the aggregate credit-to-savings ratio. This is one-third of China's overall credit boom and over one-half of what is unexplained by the government's stimulus package.

6 Conclusion

This paper has explored the unintended consequences of a liquidity regulation and the dynamics of China's shadow banking sector. We argued that shadow banking arose among small and medium-sized banks to evade stricter liquidity rules imposed by Chinese regulators. We also argued that shadow banking competes with China's Big Four banks, prompting the Big Four to use their interbank market power to undermine the shadow banks. The end

[39] The calibrated parameters are: $\omega = 126.84$, $\delta_1 = 266.36$, $\delta_2 = 0.374$, $\phi_k = 0.0335$, and $\overline{\theta} = 0.1325$.

result has been an increase in total credit and an increase in interbank rates, making the stricter liquidity rules entirely counter-productive.

References

Acharya, V., P. Schnabl, and G. Suarez. 2013. "Securitization Without Risk Transfer." *Journal of Financial Economics*, 107(3), pp. 515-536.

Adrian, T., A. Ashcraft, H. Boesky, and Z. Pozsar. 2012. "Shadow Banking." Federal Reserve Bank of New York Staff Reports, no. 458.

Adrian, T., A. Ashcraft, and N. Cetorelli. 2013. "Shadow Bank Monitoring." Federal Reserve Bank of New York Staff Reports, no. 638.

Allen, F., Y. Qian, G. Tu, and F. Yu. 2015. "Entrusted Loans: A Close Look at China's Shadow Banking System." Mimeo.

Brandt, L., J. Van Biesebroeck, and Y. Zhang. 2012. "Creative Accounting or Creative Destruction? Firm-Level Productivity Growth in Chinese Manufacturing." *Journal of Development Economics*, 97(2), pp. 339-351.

Brunnermeier, M. 2009. "Deciphering the Liquidity and Credit Crunch 2007-2008." *Journal of Economic Perspectives*, 23(1), pp. 77-100.

Cetorelli, N. and S. Peristiani. 2012. "The Role of Banks in Asset Securitization." FRBNY Economic Policy Review, July 2012, pp. 47-63.

Corbae, D. and P. D'Erasmo. 2013. "A Quantitative Model of Banking Industry Dynamics." Mimeo.

Covitz, D., N. Liang, and G. Suarez. 2013. "The Evolution of a Financial Crisis: Collapse of the Asset-Backed Commercial Paper Market." *Journal of Finance*, 68(3), pp. 815-848.

Diamond, D. and P. Dybvig. 1983. "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy*, 91(3), pp. 401-419.

Diamond, D. and A. Kashyap. 2015. "Liquidity Requirements, Liquidity Choice, and Financial Stability." In preparation for *The Handbook of Macroeconomics, Volume 2* (eds. J. Taylor and H. Uhlig).

Dobson, W. and A. Kashyap. 2006. "The Contradiction in China's Gradualist Banking Reforms." *Brookings Papers on Economic Activity*, Fall 2006, pp. 103-148.

Elliott, D., A. Kroeber, and Y. Qiao. 2015. "Shadow Banking in China: A Primer." Brookings Economic Studies.

Farhi, E., M. Golosov, and A. Tsyvinski. 2009. "A Theory of Liquidity and Regulation of Financial Intermediation." *Review of Economic Studies*, 76(3), pp. 973-992.

Financial Stability Board (FSB). 2011. "Shadow Banking: Scoping the Issues." April 12.

Gorton, G. and A. Metrick. 2012. "Securitized Banking and the Run on Repo." *Journal of Financial Economics*, 104(3), pp. 425-451.

Hsieh, C. and P. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4), pp. 1403-1448.

Hsieh, C. and Z. Song. 2015. "Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China." *Brookings Papers on Economic Activity*, forthcoming.

Kacperczyk, M. and P. Schnabl. 2013. "How Safe are Money Market Funds?" *Quarterly Journal of Economics*, 128(3), pp. 1073-1122.

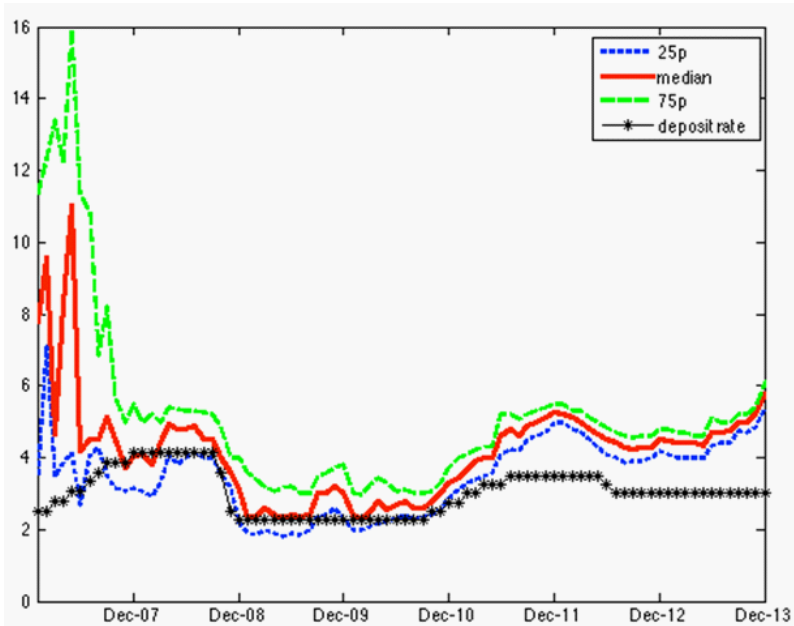
Krishnamurthy, A., S. Nagel, and D. Orlov. 2014. "Sizing Up Repo." *Journal of Finance*, 69(6), pp. 2381-2417.

Lardy, N. 2014. *Markets over Mao: The Rise of Private Business in China.* Peterson Institute for International Economics.

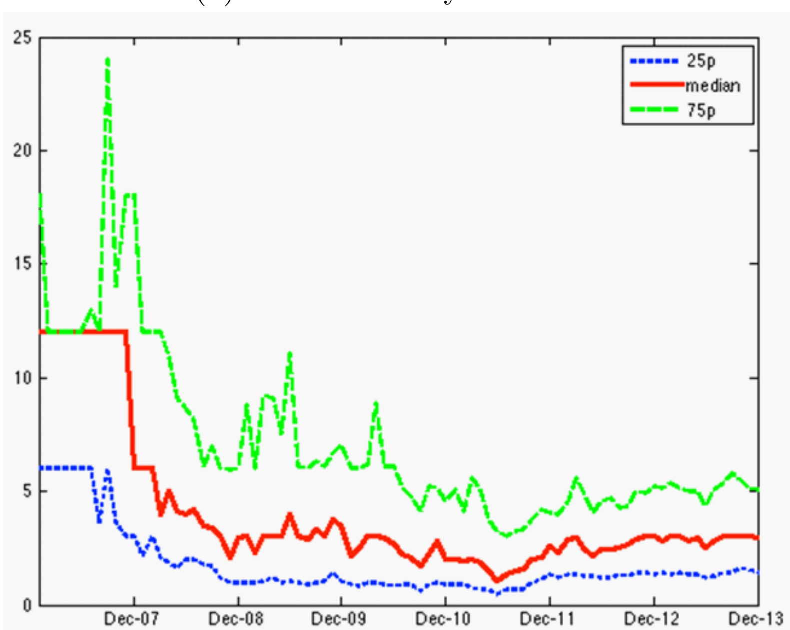
Song, Z., K. Storesletten, and F. Zilibotti. 2011. "Growing Like China." *American Economic Review*, 101(1), pp. 202-241.

Figure 1

(a) Annualized WMP Returns (%, Maturity $\leq$ 1Yr)



(b) WMP Maturity in Months



Source: Wind Financial Terminal

Figure 2

Anatomy of a Wealth Management Product

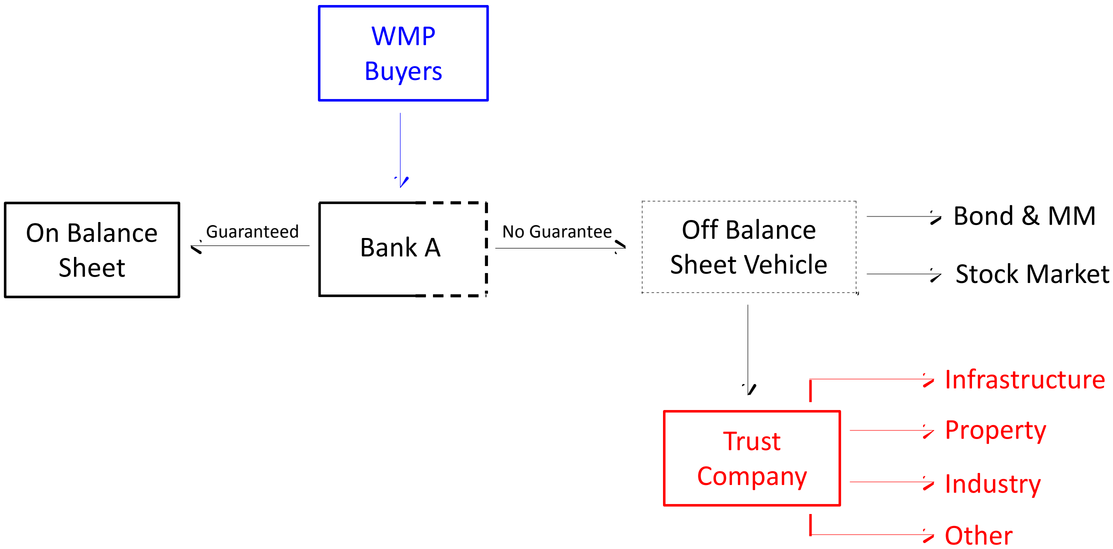
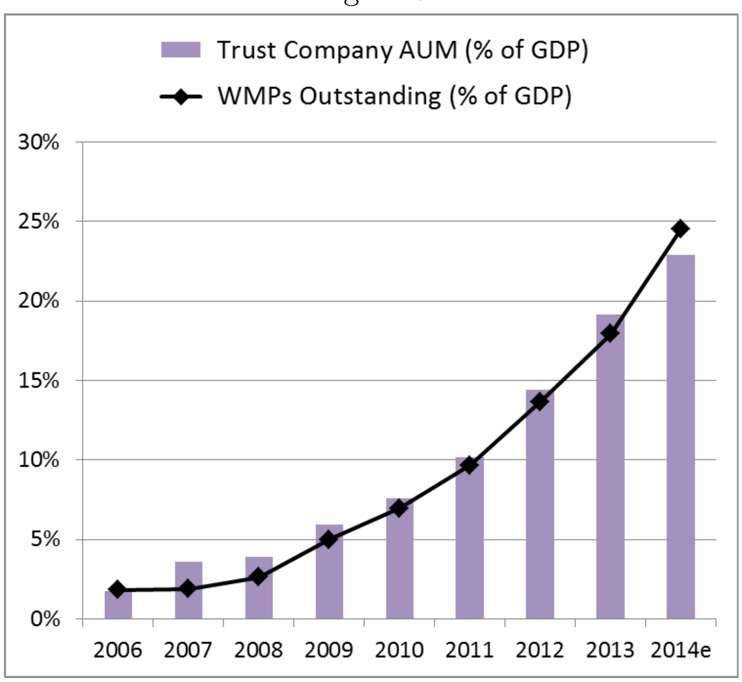
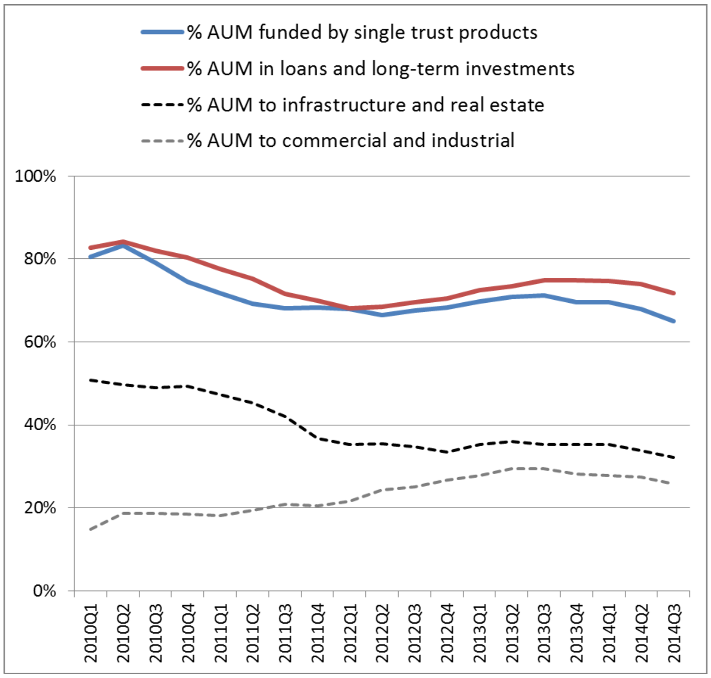


Figure 3



Source: PBOC, CBRC, IMF, China Trustee Association, KPMG China Trust Surveys

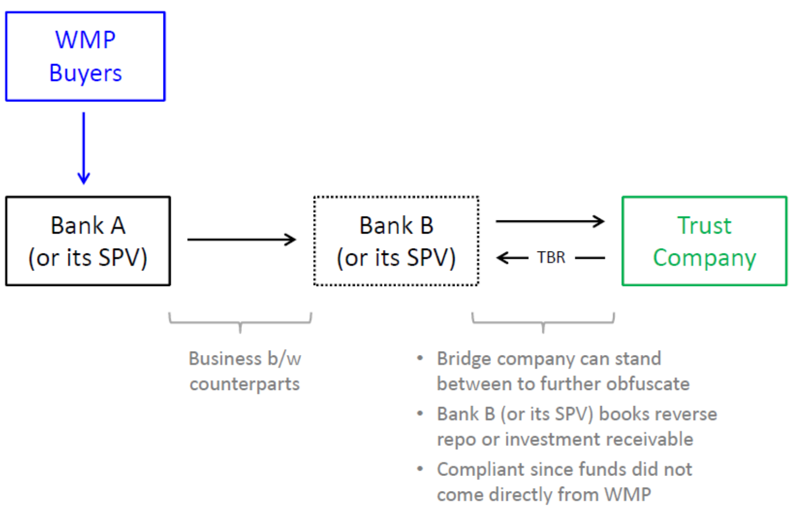
Figure 4



Source: China Trustee Association

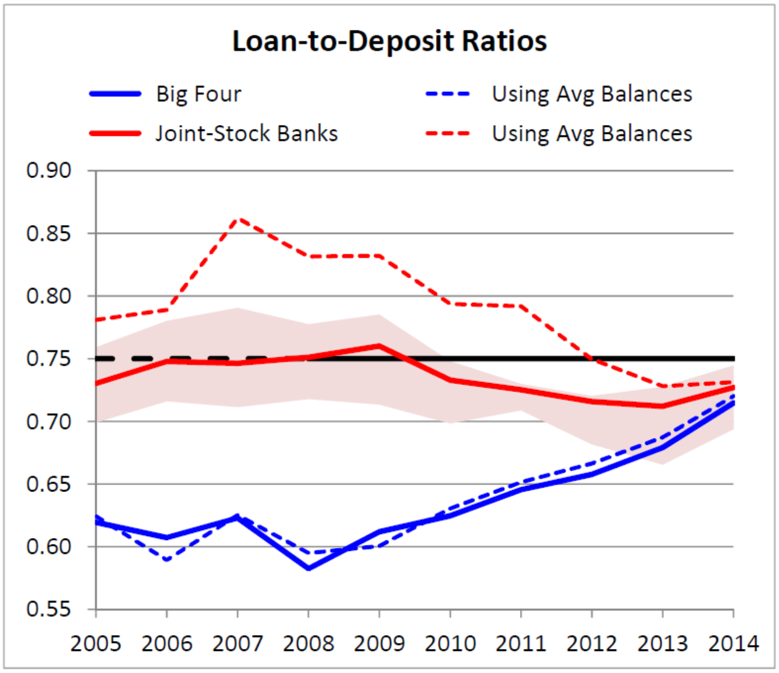
Figure 5

Business with Counterparts



Notes: TBR stands for trust beneficiary right; SPV is an off-balance-sheet vehicle

Figure 6

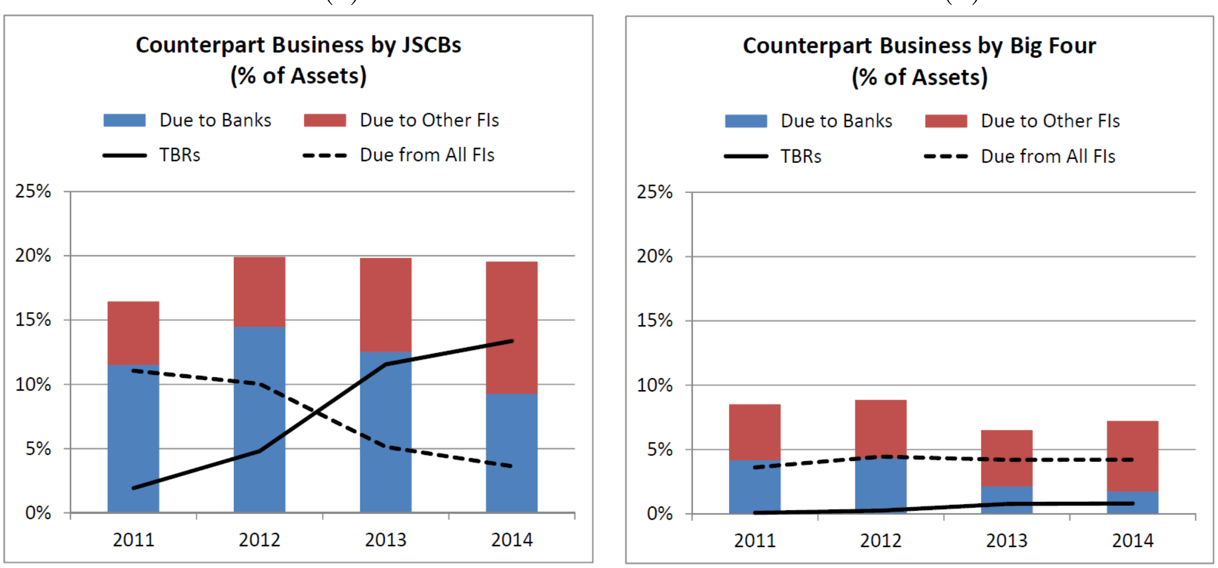


Source: Bankscope and bank annual reports. Shaded area is interquartile range.

Figure 7

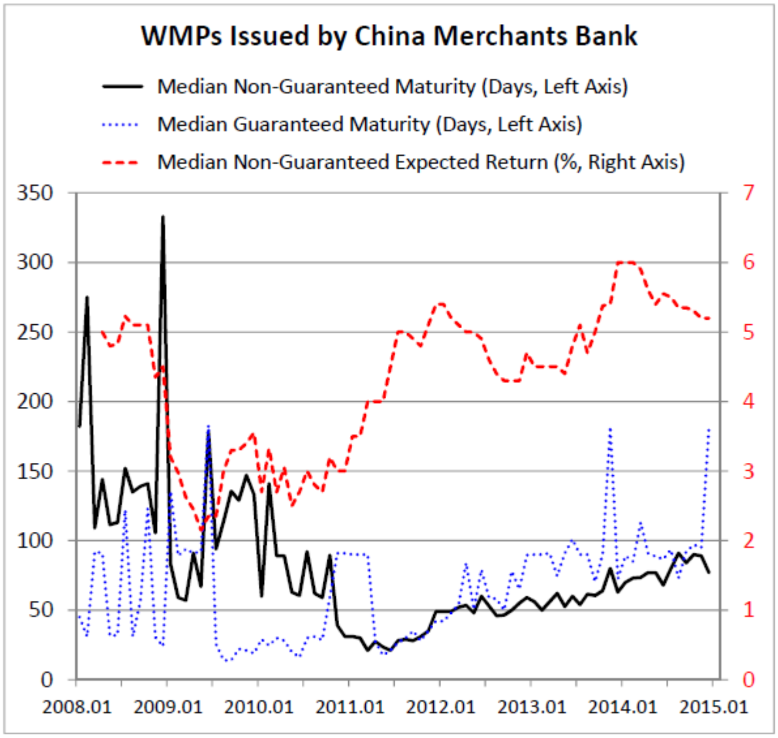
Panel (a)

Panel (b)



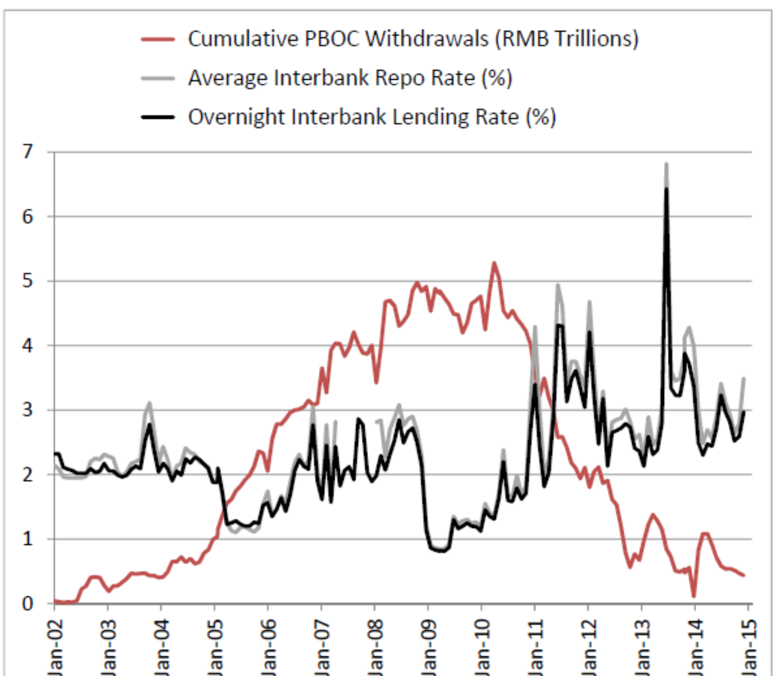
Source: Bank annual reports. The graphs report domestic balances only.

Figure 8



Source: Wind Financial Terminal

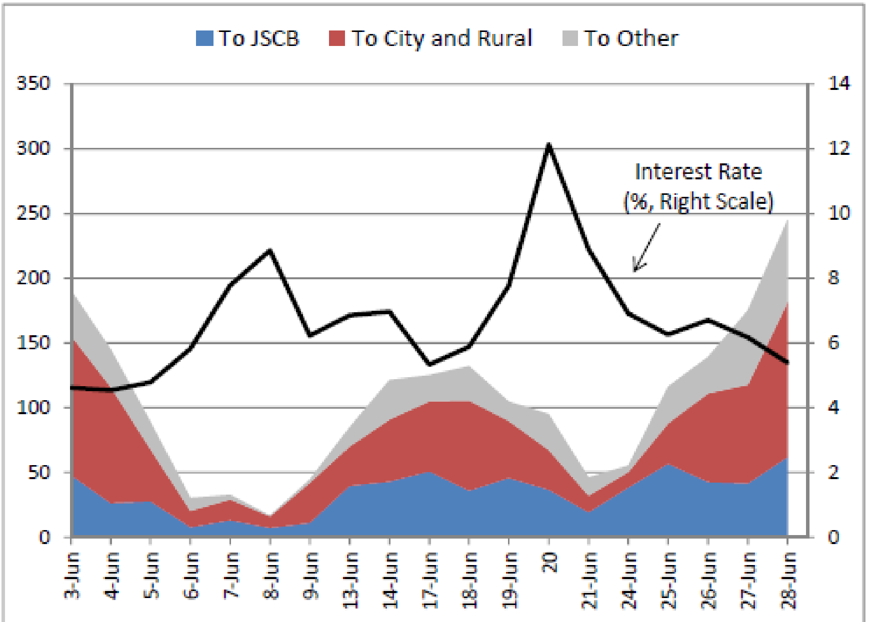
Figure 9



Source: PBOC and Wind Financial Terminal

Figure 10

(a) Repo Lending by Big Banks (RMB Billions)



Note: Excludes lending between big banks

(b) Repo Lending by Policy Banks (RMB Billions)

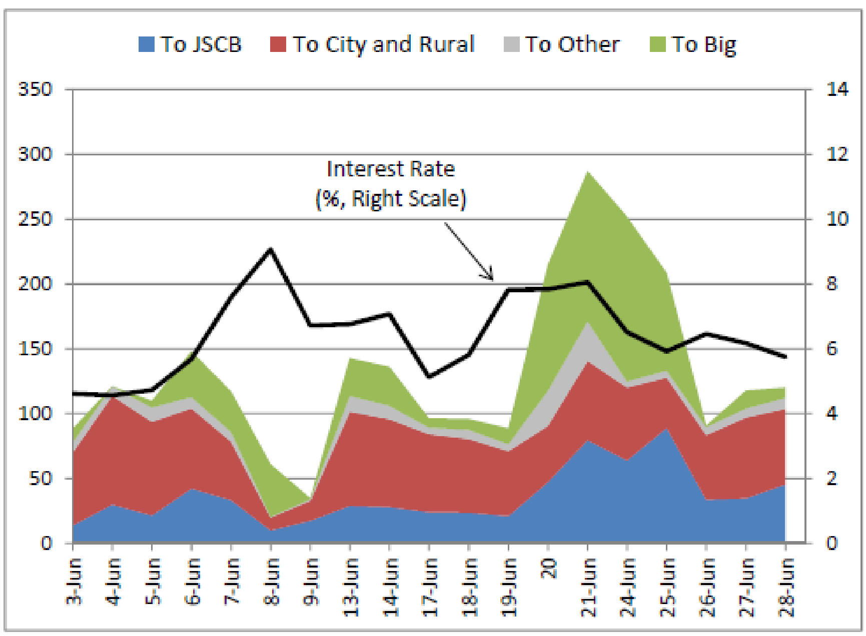


Figure 11

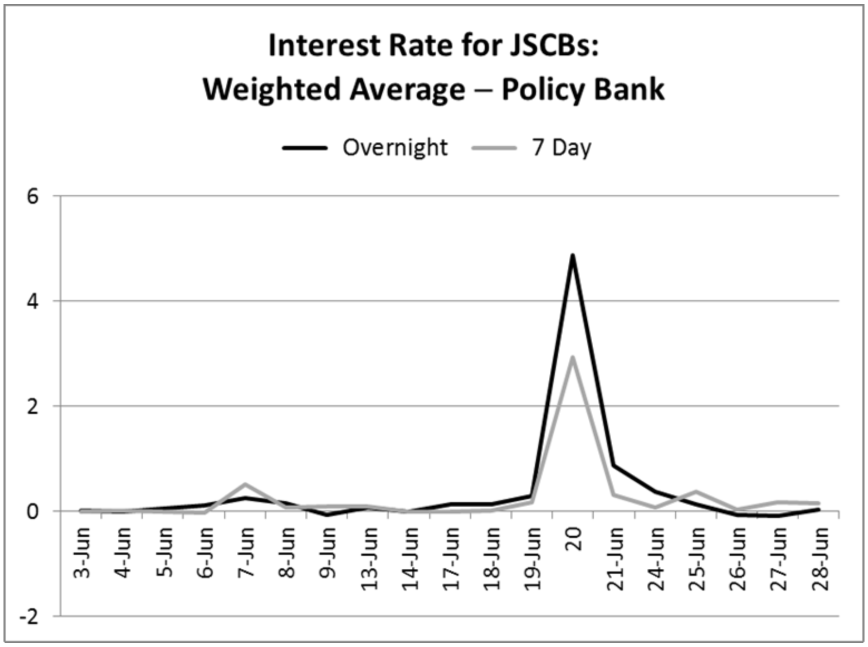


Figure 12

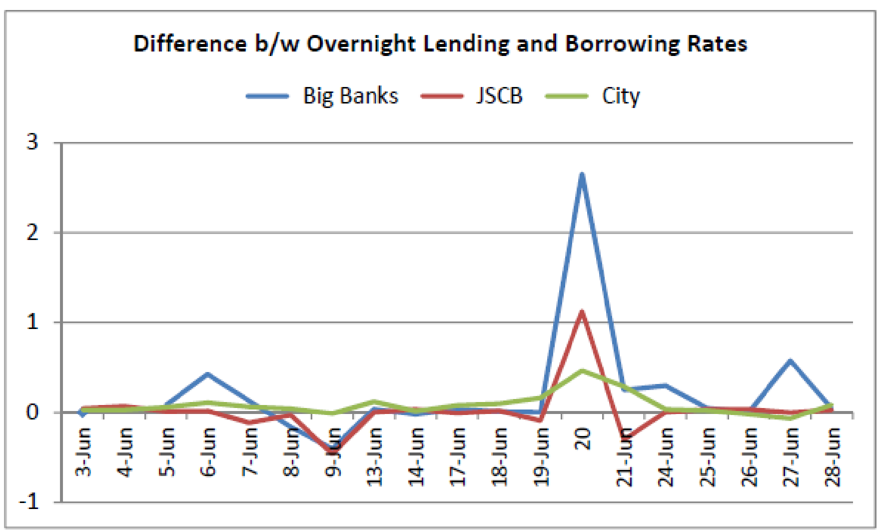


Table 1

Granger Causality Wald Tests

	χ^2	Prob > χ^2
H0: WMP batches of SMBs do not cause WMP batches of Big Four	21.104	**0.002**
H0: WMP batches of Big Four do not cause WMP batches of SMBs	5.5264	**0.478**

Notes: We use detrended monthly data from Wind and estimate VARs with six lags

Table 2

Calibration Results

	(1) Model $\alpha = 0.14$	(2) Data 2007	(3) Model $\alpha = 0.25$	(4) Data 2014
Average Interbank Rate $(\pi i_L^\ell + (1-\pi)\, i_L^h)$	3.35%	3.1%	3.6%	3.6%
Small Bank WMPs (W_j)	0.03	NA	0.10	0.10
Big Bank WMPs (W_k)	0.01	NA	0.05	0.05
Big Bank Funding Share (X_k)	0.52	0.55	0.45	0.45
Big Bank Loan-to-Deposit Ratio $(1 - \frac{R_k}{X_k})$	58%	62%	70%	70%
Credit-to-Savings Ratio $(1 - R_j - R_k)$	72.1%	65%	75.3%	75%

Appendix A – Proofs

Proof of Proposition 1

By contradiction. Suppose $\rho_2 > 0$. If $\mu_j > 0$, then $R_j = 0$ so equation (4) implies $i_L = \frac{\overline{\theta} X}{\psi}$. Substituting into equation (9) then implies $\xi_j > 0$ if and only if $\phi < \frac{(1+i_A)^2 - 1}{X} - \frac{\overline{\theta}^2 X}{\psi X} \equiv \overline{\phi}_1$ (where we have used $X_j = X$ in a symmetric equilibrium). If instead $\mu_j = 0$, then equation (7) implies $i_L = (1 + i_A)^2 - 1$. Substituting into equation (9) then implies $\xi_j > 0$ if and only if $\phi < \frac{1-\overline{\theta}}{X}\left[(1 + i_A)^2 - 1\right] \equiv \overline{\phi}_2$. Defining $\overline{\phi} \equiv \min\left\{\overline{\phi}_1, \overline{\phi}_2\right\}$ completes the proof. ■

Proof of Proposition 2

With $\rho_2 = 0$, the equilibrium is characterized by (4), (7), and:

$$\xi_j = \frac{\alpha \mu_j}{2\left(1 - \overline{\theta}\right)}$$

$$\mu_j \left[R_j - \alpha\left(X - \omega \xi_j\right)\right] = 0 \text{ with complementary slackness}$$

There is an implicit refinement here since we are writing $\xi_j = \frac{\alpha \mu_j}{2\left(1-\overline{\theta}\right)}$ instead of $\xi_j = \frac{\alpha \mu_j \tau_j}{2\left(1-\overline{\theta}\right)}$. Both produce $\xi_j = 0$ if $\alpha \mu_j = 0$ so the refinement only applies if $\alpha \mu_j > 0$. Return to equations (8) and (9) with $\rho_2 = 0$ and $\alpha \mu_j > 0$. If $\xi_j > 0$, then $\eta_j^1 > 0$. This implies $\tau_j = 1$ which confirms $\xi_j > 0$. If $\xi_j = 0$, then $\eta_j^1 = \eta_j^0$. This implies $\tau_j \in [0, 1]$. However, any $\tau_j \in (0, 1]$ would return $\xi_j > 0$, violating $\xi_j = 0$. We thus eliminate $\xi_j = 0$ by refinement. Instead, $\alpha \mu_j > 0$ is associated with $\xi_j > 0$ and thus $\tau_j = 1$. For this reason, we write $\xi_j = \frac{\alpha \mu_j}{2\left(1-\overline{\theta}\right)}$. We can now proceed with the rest of the proof. There are two cases:

1. If $\mu_j = 0$, then $\xi_j = 0$ and $1 + i_L = (1 + i_A)^2$. Equation (4) then pins down R_j. To ensure that $R_j \geq \alpha\left(X - \omega \xi_j\right)$ is satisfied, we need $\alpha \leq \overline{\theta} - \frac{\psi\left[(1+i_A)^2 - 1\right]}{X} \equiv \widetilde{\alpha}$. We have now established $\xi_j = 0$ if $\alpha \leq \widetilde{\alpha}$.

2. If $\mu_j > 0$, then complementary slackness implies $R_j = \alpha\left(X - \omega \xi_j\right)$. Combining with the other equilibrium conditions, we find that $\mu_j > 0$ delivers:

$$i_L = \frac{\alpha^2 \omega\left[(1 + i_A)^2 - 1\right] - 2\left(1 - \overline{\theta}\right)\left(\alpha - \overline{\theta}\right) X}{\alpha^2 \omega + 2\psi\left(1 - \overline{\theta}\right)} \tag{21}$$

Verifying $\mu_j > 0$ is equivalent to verifying $1 + i_L < (1 + i_A)^2$. This reduces to $\alpha > \widetilde{\alpha}$. If $\widetilde{\alpha} \geq 0$, then we have established $\xi_j > 0$ with $\tau_j = 1$ for any $\alpha > \widetilde{\alpha}$.

Defining $\overline{\alpha} = \max\{\widetilde{\alpha}, 0\}$ completes the proof. ■

Proof of Proposition 3

Consider $\alpha = 0$. If $\mu_j = 0$, then (7) implies $i_L = (1 + i_A)^2 - 1$ which is the highest feasible interbank rate. If instead $\mu_j > 0$, then the liquidity rule binds. In particular, $R_j = \alpha (X_j - \tau_j W_j)$ which is just $R_j = 0$ when $\alpha = 0$. We can then conclude $i_L = \frac{\overline{\theta} X}{\psi}$ from (4). Note that $\mu_j > 0$ is verified if and only if $\frac{\overline{\theta} X}{\psi} < (1 + i_A)^2 - 1$.

Based on the results so far, we can see that the interbank rate at $\alpha = 0$ is independent of ρ_2. Let i_{L0} denote the interbank rate at $\alpha = 0$ and let $i_{L1}(\rho_2)$ denote the interbank rate at some $\alpha > 0$. From (4), we know $i_{L1}(\rho_2) = \frac{\overline{\theta} X}{\psi} - \frac{R_{j1}(\rho_2)}{\psi}$, where $R_{j1}(\rho_2)$ is reserve holdings at the $\alpha > 0$ being considered. The rest of the proof proceeds by contradiction. In particular, suppose $i_{L1}(\rho_2) > i_{L0}$. Then $\alpha = 0$ must be associated with $\mu_j > 0$, otherwise i_{L0} would be the highest feasible interbank rate and the supposition would be incorrect. We can thus write $i_{L0} = \frac{\overline{\theta} X}{\psi}$ and $i_{L1}(\rho_2) = i_{L0} - \frac{R_{j1}(\rho_2)}{\psi}$. The only way to get $i_{L1}(\rho_2) > i_{L0}$ is then $R_{j1}(\rho_2) < 0$ which is impossible. ■

Proof of Proposition 4

Start with general α. The derivatives of the big bank's objective function are:

$$\frac{\partial \Upsilon_k}{\partial \xi_k} \propto -\frac{2\omega\left(1 - \overline{\theta}\right)}{1 - \pi}\xi_k - \left[\frac{(1 + i_A)^2 - 1}{1 - \pi} - i_L^h\right]\frac{\partial R_k}{\partial \xi_k} + \left[\frac{(1 + i_A)^2 - 1 - \phi X_k}{1 - \pi} - \theta_h i_L^h\right]\frac{\partial X_k}{\partial \xi_k}$$

$$\frac{\partial \Upsilon_k}{\partial i_L^h} \propto R_k - \theta_h X_k - \left[\frac{(1 + i_A)^2 - 1}{1 - \pi} - i_L^h\right]\frac{\partial R_k}{\partial i_L^h} + \left[\frac{(1 + i_A)^2 - 1 - \phi X_k}{1 - \pi} - \theta_h i_L^h\right]\frac{\partial X_k}{\partial i_L^h}$$

It will be convenient to reduce these derivatives to a core set of variables (ξ_j, ξ_k, and i_L^h). If $\mu_j > 0$, then the complementary slackness in equation (13) implies:

$$R_j = \alpha\left(X_j - \omega\xi_j\right) \tag{22}$$

With $\delta_1 + \delta_2 = 0$ and $\overline{\xi_j} = \xi_j$, equations (10) and (11) are:

$$X_j = 1 - \delta_0 + \delta_1\left(\xi_j - \xi_k\right) \tag{23}$$

$$X_k = \delta_0 + \delta_1\left(\xi_k - \xi_j\right) \tag{24}$$

Substitute (22) to (24) into equation (12) to write:

$$R_k = \delta_0\theta_h + (1-\delta_0)\left(\overline{\theta}-\alpha\right) + \delta_1\left(\theta_h - \overline{\theta} + \alpha\right)\left(\xi_k - \xi_j\right) + \alpha\omega\xi_j - \psi i_L^h \tag{25}$$

Finally, combine equations (14) and (15) to get:

$$\xi_j = \frac{\alpha(1-\pi)}{2\left(1-\overline{\theta}\right)}\left[\frac{(1+i_A)^2-1}{1-\pi} - i_L^h\right] \tag{26}$$

We can now write $\frac{\partial \Upsilon_k}{\partial \xi_k} = 0$ as:

$$\xi_k = \frac{\delta_1\left[\left(1-\theta_h+\overline{\theta}-\alpha\right)\left[(1+i_A)^2-1\right] - \phi\delta_0 + \phi\delta_1\xi_j - \left(\overline{\theta}-\alpha\right)(1-\pi)\, i_L^h\right]}{2\omega\left(1-\overline{\theta}\right)+\phi\delta_1^2} \tag{27}$$

We can also write $\frac{\partial \Upsilon_k}{\partial i_L^h} = 0$ as:

$$\begin{aligned} i_L^h \quad = \quad & \frac{\left[\frac{\psi}{1-\pi} + \frac{\alpha\left[\alpha\omega+\delta_1\left(1-\theta_h+\overline{\theta}-\alpha\right)\right]}{2\left(1-\overline{\theta}\right)}\right]\left[(1+i_A)^2-1\right]}{2\psi + \frac{\alpha(1-\pi)}{2\left(1-\overline{\theta}\right)}\left[\alpha\omega+\delta_1\left(\overline{\theta}-\alpha\right)\right]} \\ & + \frac{(1-\delta_0)\left(\overline{\theta}-\alpha\right) - \frac{\alpha\phi\delta_0\delta_1}{2\left(1-\overline{\theta}\right)} + \alpha\omega\xi_j - \delta_1\left[\overline{\theta}-\alpha+\frac{\alpha\phi\delta_1}{2\left(1-\overline{\theta}\right)}\right]\left(\xi_k-\xi_j\right)}{2\psi + \frac{\alpha(1-\pi)}{2\left(1-\overline{\theta}\right)}\left[\alpha\omega+\delta_1\left(\overline{\theta}-\alpha\right)\right]} \end{aligned} \tag{28}$$

Remark 1 *As long as the big bank's inequality constraints are non-binding, the equilibrium is a triple* $\left\{\xi_j, \xi_k, i_L^h\right\}$ *that solves (26), (27), and (28). We must therefore check that the solution to these equations satisfies* $\xi_k \geq 0$ *along with* $R_k > \alpha X_k$ *and* $\mu_j > 0$*. We also need to check* $W_j \leq X_j$ *and* $W_k \leq X_k$ *so that deposits are non-negative. Finally, we want to check that* $i_L^\ell = 0$ *does not result in a liquidity shortage when the big bank realizes* θ_ℓ *at* $t = 1$*.*

The rest of this proof focuses on $\alpha = 0$. Notice $\xi_j = 0$ from (26). As discussed in the main text, we also want $\xi_k = 0$. Subbing $\alpha = 0$ and $\xi_j = \xi_k = 0$ into (27) and (28) yields:

$$\delta_1\left[\frac{\left(1-\theta_h+\overline{\theta}\right)\left[(1+i_A)^2-1\right]-\phi\delta_0}{\overline{\theta}(1-\pi)} - i_L^h\right] = 0 \tag{29}$$

$$i_L^h = \frac{(1+i_A)^2-1}{2(1-\pi)} + \frac{\overline{\theta}(1-\delta_0)}{2\psi} \tag{30}$$

To verify $\xi_k = 0$, we must verify that (29) holds when i_L^h is given by (30). This requires either $\delta_1 = 0$ or:

$$\phi = \frac{1}{\delta_0}\left[1 - \theta_h + \frac{\overline{\theta}}{2}\right]\left[(1+i_A)^2 - 1\right] - \frac{\overline{\theta}^2(1-\pi)(1-\delta_0)}{2\psi\delta_0} \equiv \phi^* \tag{31}$$

In other words, we can use either $\delta_1 = 0$ or the combination of $\delta_1 > 0$ and $\phi = \phi^*$ to get ξ_k exactly zero at $\alpha = 0$. Note that $W_j \leq X_j$ and $W_k \leq X_k$ are trivially true with $\xi_j = \xi_k = 0$. We now need to check $R_k > \alpha X_k$ and $\mu_j > 0$. Using (14) and (30), rewrite $\mu_j > 0$ as:

$$\frac{(1+i_A)^2 - 1}{1-\pi} > \frac{\overline{\theta}(1-\delta_0)}{\psi} \tag{32}$$

Note that condition (32) is also sufficient for $\phi^* > 0$. With $\mu_j > 0$ verified, we can substitute $\alpha = 0$ into equation (22) to get $R_j = 0$. The next step is to check $R_k > \alpha X_k$ which is simply $R_k > 0$ at $\alpha = 0$. Recall that R_k is given by equation (25). Use $\alpha = 0$ and $\xi_j = \xi_k = 0$ along with i_L^h as per (30) to rewrite equation (25) as:

$$R_k = \theta_h\delta_0 + \frac{\overline{\theta}(1-\delta_0)}{2} - \psi\frac{(1+i_A)^2 - 1}{2(1-\pi)} \tag{33}$$

The condition for $R_k > 0$ is therefore:

$$\frac{(1+i_A)^2 - 1}{1-\pi} < \frac{\overline{\theta}(1-\delta_0)}{\psi} + \frac{2\delta_0\theta_h}{\psi} \tag{34}$$

The last step is to check that there is sufficient liquidity at $t = 1$ when the big bank's liquidity shock is low. The demand for liquidity in this case will be $\overline{\theta}X_j + \theta_\ell X_k$. The supply of liquidity will be $R_j + R_k$ since we have fixed $i_L^\ell = 0$. We already know $\xi_j = \xi_k = 0$ at $\alpha = 0$. Therefore, $X_j = 1 - \delta_0$ and $X_k = \delta_0$. We also know $R_j = 0$ and R_k as per (33). Therefore, $R_j + R_k \geq \overline{\theta}X_j + \theta_\ell X_k$ can be rewritten as:

$$\frac{(1+i_A)^2 - 1}{1-\pi} \leq \frac{2\delta_0(\theta_h - \theta_\ell)}{\psi} - \frac{\overline{\theta}(1-\delta_0)}{\psi} \tag{35}$$

Condition (35) is stricter than (34) so we can drop (34). We now just need to make sure that conditions (32) and (35) are not mutually exclusive. Using $\overline{\theta} \equiv \pi\theta_\ell + (1-\pi)\theta_h$, this requires:

$$\theta_\ell < \left[1 - \frac{1-\delta_0}{\delta_0 + \pi(1-\delta_0)}\right]\theta_h \tag{36}$$

The right-hand side of (36) is positive if and only if:

$$\pi > \frac{1-2\delta_0}{1-\delta_0} \tag{37}$$

Therefore, with θ_ℓ sufficiently low and π sufficiently high, conditions (32) and (35) define a non-empty interval for i_A, completing the proof. ■

Proof of Proposition 5

Fixed Funding Share Impose $\alpha = \overline{\theta}$ and $\delta_1 = 0$ on equations (26), (27), and (28). The resulting system can be written as $\xi_k = 0$ and:

$$\xi_j = \frac{\frac{\overline{\theta}\psi}{2}\left[(1+i_A)^2 - 1\right]}{2\psi\left(1-\overline{\theta}\right) + \omega\overline{\theta}^2(1-\pi)} \tag{38}$$

$$i_L^h = \frac{\left[\frac{\psi\left(1-\overline{\theta}\right)}{1-\pi} + \omega\overline{\theta}^2\right]\left[(1+i_A)^2 - 1\right]}{2\psi\left(1-\overline{\theta}\right) + \omega\overline{\theta}^2(1-\pi)} \tag{39}$$

With $\delta_1 = 0$ in equations (23) and (24), the funding shares are $X_j = 1-\delta_0$ and $X_k = \delta_0$. Impose along with $\alpha = \overline{\theta}$ on equations (22) and (25) to get:

$$R_k = \theta_h\delta_0 + \omega\overline{\theta}\xi_j - \psi i_L^h$$

$$R_j + R_k = \overline{\theta}(1-\delta_0) + \theta_h\delta_0 - \psi i_L^h$$

where ξ_j and i_L^h are given by (38) and (39) respectively. We now need to go through all the steps in Remark 1 to establish the equilibrium for $\alpha = \overline{\theta}$ and fixed funding shares. Using equations (14) and (39), we can see that $\mu_j > 0$ is trivially true. Using $\xi_k = 0$ and $X_k = \delta_0$, we can also see that $W_k \leq X_k$ is trivially true. The condition for $W_j \leq X_j$ is:

$$\frac{(1+i_A)^2 - 1}{1-\pi} \leq \frac{2(1-\delta_0)}{\psi}\left[\overline{\theta} + \frac{2\psi\left(1-\overline{\theta}\right)}{\omega\overline{\theta}(1-\pi)}\right] \tag{40}$$

The conditions for $R_k > \overline{\theta}X_k$ and $R_j + R_k \geq \overline{\theta}X_j + \theta_\ell X_k$ are respectively:

$$\frac{(1+i_A)^2 - 1}{1-\pi} < \frac{2\pi(\theta_h - \theta_\ell)\delta_0}{\psi} \tag{41}$$

$$\frac{\left(1+i_A\right)^2-1}{1-\pi} \leq \frac{2\psi\left(1-\overline{\theta}\right)+\omega\overline{\theta}^2\left(1-\pi\right)}{\psi\left(1-\overline{\theta}\right)+\omega\overline{\theta}^2\left(1-\pi\right)} \frac{\left(\theta_h-\theta_\ell\right)\delta_0}{\psi} \tag{42}$$

Now, for the interbank rate to increase when moving from $\alpha = 0$ to $\alpha = \overline{\theta}$, we need (39) to exceed (30). Equivalently, we need:

$$\frac{\left(1+i_A\right)^2-1}{1-\pi} > \frac{\overline{\theta}\left(1-\delta_0\right)}{\psi}\left[1+\frac{2\psi\left(1-\overline{\theta}\right)}{\omega\overline{\theta}^2\left(1-\pi\right)}\right] \tag{43}$$

We must now collect all the conditions involved in the $\alpha = 0$ and $\alpha = \overline{\theta}$ equilibria and make sure they are mutually consistent. There are two lowerbounds on i_A, namely (32) and (43). Condition (43) is clearly stricter so it is the relevant lowerbound. There are also four upperbounds on i_A, namely (35), (40), (41), and (42). For the lowerbound in (43) to not violate any of these upperbounds, we need:

$$\frac{\psi\left(1-\overline{\theta}\right)}{\omega\left(1-\pi\right)} < \overline{\theta}^2 \min\left\{\frac{\pi\left(\theta_h-\theta_\ell\right)\delta_0}{\overline{\theta}\left(1-\delta_0\right)}-\frac{1}{2}, \frac{\left(\theta_h-\theta_\ell\right)\delta_0}{\overline{\theta}\left(1-\delta_0\right)}-1\right\}$$

This inequality is only possible if the right-hand side is positive. Therefore, we need:

$$\theta_\ell < \left[1-\frac{1-\delta_0}{\min\left\{\delta_0+\pi\left(1-\delta_0\right),\pi\left(1+\delta_0\right)\right\}}\right]\theta_h \tag{44}$$

Once again, the right-hand side must be positive so we need:

$$\pi > \max\left\{\frac{1-2\delta_0}{1-\delta_0}, \frac{1-\delta_0}{1+\delta_0}\right\} \tag{45}$$

Notice that (44) and (45) are just refinements of (36) and (37). We can now conclude that the model with fixed funding shares generates the desired results under the following conditions: π sufficiently high, θ_ℓ and $\frac{\psi}{\omega}$ sufficiently low, and i_A within an intermediate range. □

Endogenous Funding Share Return to equations (26), (27), and (28). Impose $\alpha = \overline{\theta}$ and $\delta_1 = \omega$ with $\phi = \phi^*$ as per (31). Combine to get:

$$i_L^h = \frac{\left(1+i_A\right)^2-1}{1-\pi} - \frac{\left[\frac{2\psi}{1-\pi}+\frac{\omega\overline{\theta}^2}{2\left(1-\overline{\theta}\right)+\phi^*\omega}\right]\frac{\left(1+i_A\right)^2-1}{2\left(1-\pi\right)}-\frac{\omega\overline{\theta}^3\left(1-\delta_0\right)}{2\psi\left[2\left(1-\overline{\theta}\right)+\phi^*\omega\right]}}{\frac{2\psi}{1-\pi}+\frac{\omega\overline{\theta}^2}{2\left(1-\overline{\theta}\right)}\left[2+\frac{\phi^*\omega}{2\left(1-\overline{\theta}\right)+\phi^*\omega}\right]} \tag{46}$$

$$\xi_k = \frac{\frac{\overline{\theta}(1-\pi)}{2}\left[\frac{\overline{\theta}(1-\delta_0)}{\psi} + \left(\frac{\phi^*\omega}{1-\overline{\theta}} - 1\right)\frac{(1+i_A)^2-1}{1-\pi} - \frac{\phi^*\omega}{1-\overline{\theta}} i_L^h\right]}{2\left(1-\overline{\theta}\right) + \phi^*\omega} \tag{47}$$

We now need to go through the steps in Remark 1 to establish the equilibrium for $\alpha = \overline{\theta}$ and endogenous funding shares. The expressions here are more complicated so we proceed by finding one value of i_A that satisfies all the steps in Remark 1. A continuity argument will then allow us to conclude that all the steps are satisfied for a non-empty range of i_A.

Consider i_A such that:

$$\frac{(1+i_A)^2 - 1}{1-\pi} = \frac{\overline{\theta}}{\psi} \tag{48}$$

Substituting into (31) then pins down ϕ^* as:

$$\phi^* = \frac{\overline{\theta}\,(1-\pi)}{\psi}\left[\frac{1-\theta_h}{\delta_0} + \frac{\overline{\theta}}{2}\right] \tag{49}$$

From the proof of Proposition 4, we already have (32) and (35) as restrictions on i_A. We also have (36) as an upperbound on θ_ℓ and (37) as a lowerbound on π. It is easy to see that i_A as defined in (48) satisfies (32). For (48) to also satisfy (35), we need:

$$\theta_\ell < \left[1 - \frac{2-\delta_0}{2\delta_0 + \pi\,(2-\delta_0)}\right]\theta_h \tag{50}$$

$$\pi > \frac{2-3\delta_0}{2-\delta_0} \tag{51}$$

Conditions (50) and (51) are stricter than (36) and (37). We can thus drop (36) and (37).

The first step is to verify $\mu_j > 0$. Use (14) and (46) to write $\mu_j > 0$ as:

$$\frac{(1+i_A)^2-1}{1-\pi}\left[1 + \frac{2\psi\left[2\left(1-\overline{\theta}\right)+\phi^*\omega\right]}{\omega\overline{\theta}^2\,(1-\pi)}\right] > \frac{\overline{\theta}\,(1-\delta_0)}{\psi}$$

This is true by condition (32).

The second step is to verify $\xi_k > 0$. Substituting (46) into (47), we see that we need:

$$\frac{(1+i_A)^2-1}{1-\pi}\left[1 - \frac{\phi^*}{\frac{2\left(1-\overline{\theta}\right)}{\omega} + \frac{\overline{\theta}^2(1-\pi)}{\psi}}\right] < \frac{\overline{\theta}\,(1-\delta_0)}{\psi} \tag{52}$$

Using i_A as per (48) and ϕ^* as per (49):

$$\frac{\psi\left(1-\overline{\theta}\right)}{\omega\left(1-\pi\right)} < \frac{\overline{\theta}}{2\delta_0^2}\underbrace{\left[1-\theta_h-\overline{\theta}\delta_0\left(\delta_0-\frac{1}{2}\right)\right]}_{\text{call this } Z_1} \tag{53}$$

If $Z_1 > 0$, then (53) requires $\frac{\psi}{\omega}$ sufficiently low. Note that $Z_1 > 0$ can be made true for any $\delta_0 \in (0,1)$ by assuming $\overline{\theta} < 2\left(1-\theta_h\right)$ or, equivalently, $\theta_\ell < \frac{2-(3-\pi)\theta_h}{\pi}$. This is another positive ceiling on θ_ℓ provided $\pi > 3 - \frac{2}{\theta_h}$.

The third step is to verify $R_k > \overline{\theta}X_k$. Use $\alpha = \overline{\theta}$ and $\delta_1 = \omega$ to rewrite (24) and (25) as:

$$X_k = \delta_0 + \omega\left(\xi_k - \xi_j\right) \tag{54}$$

$$R_k = \delta_0\theta_h + \omega\theta_h\xi_k - \omega\left(\theta_h - \overline{\theta}\right)\xi_j - \psi i_L^h \tag{55}$$

Therefore, $R_k > \overline{\theta}X_k$ requires:

$$i_L^h < \frac{\delta_0\left(\theta_h-\overline{\theta}\right)}{\psi} + \frac{\omega\left(\theta_h-\overline{\theta}\right)}{\psi}\left(\xi_k-\xi_j\right) + \frac{\omega\overline{\theta}}{\psi}\xi_j$$

Use (47) to replace ξ_k and (26) with $\alpha = \overline{\theta}$ to replace ξ_j:

$$\begin{aligned}
&\left[1+\frac{\omega\overline{\theta}\left(1-\pi\right)}{2\psi\left(1-\overline{\theta}\right)}\left[\overline{\theta}-\frac{2\left(1-\overline{\theta}\right)\left(\theta_h-\overline{\theta}\right)}{2\left(1-\overline{\theta}\right)+\phi^*\omega}\right]\right]i_L^h \\
< \quad & \frac{\theta_h-\overline{\theta}}{\psi}\left[\delta_0+\frac{\omega\overline{\theta}^2\left(1-\pi\right)\left(1-\delta_0\right)}{2\psi\left[2\left(1-\overline{\theta}\right)+\phi^*\omega\right]}\right] - \frac{\omega\overline{\theta}\left[\left(1+i_A\right)^2-1\right]}{2\psi}\left[\frac{3\left(\theta_h-\overline{\theta}\right)}{2\left(1-\overline{\theta}\right)+\phi^*\omega}-\frac{\overline{\theta}}{1-\overline{\theta}}\right]
\end{aligned}$$

Now use (46) to replace i_L^h and rearrange to isolate i_A:

$$\begin{aligned}
&\frac{\left(1+i_A\right)^2-1}{1-\pi}\left[2\theta_h-\frac{3\overline{\theta}}{2}+\frac{2\psi\left(1-\overline{\theta}\right)}{\omega\overline{\theta}\left(1-\pi\right)}+\frac{\omega\overline{\theta}^2\left(1-\pi\right)}{4\psi\left(1-\overline{\theta}\right)}\left(3\theta_h-4\overline{\theta}\right)+\frac{\phi^*}{\overline{\theta}}\left[\frac{\omega\overline{\theta}^2}{1-\overline{\theta}}+\frac{\psi}{1-\pi}\right]\right] \\
< \quad & \left[\frac{2\delta_0\left(\theta_h-\overline{\theta}\right)}{1-\pi}\frac{2\left(1-\overline{\theta}\right)+\phi^*\omega}{\omega\overline{\theta}}-\frac{\overline{\theta}^2\left(1-\delta_0\right)}{2\psi}\right]\left[1+\frac{\overline{\theta}^2}{2}\frac{\omega\left(1-\pi\right)}{\psi\left(1-\overline{\theta}\right)}\right] \\
& + \left(\theta_h-\overline{\theta}\right)\frac{\overline{\theta}}{\psi}\left[\frac{\omega\phi^*\delta_0}{2\left(1-\overline{\theta}\right)}+\left(1-\delta_0\right)\left[1+\frac{3\overline{\theta}^2}{4}\frac{\omega\left(1-\pi\right)}{\psi\left(1-\overline{\theta}\right)}\right]\right]
\end{aligned}$$

We can simplify a bit further by using (31) to replace all instances of $\phi^*\delta_0$ then grouping like terms:

$$
\begin{aligned}
& \frac{(1+i_A)^2-1}{1-\pi}\left[\begin{array}{l} \theta_h - \frac{\overline{\theta}}{2} + \frac{2\psi\left(1-\overline{\theta}\right)}{\omega\overline{\theta}(1-\pi)} - \frac{\omega\overline{\theta}^3(1-\pi)}{4\psi\left(1-\overline{\theta}\right)} + \frac{\phi^*}{\overline{\theta}}\left[\frac{\omega\overline{\theta}^2}{1-\overline{\theta}} + \frac{\psi}{1-\pi}\right] \\ -\left(\theta_h-\overline{\theta}\right)(1-\theta_h)\left[\frac{2}{\overline{\theta}} + \frac{3\overline{\theta}}{2}\frac{\omega(1-\pi)}{\psi\left(1-\overline{\theta}\right)}\right] \end{array}\right] \\
< \; & \left[\frac{4\delta_0\left(1-\overline{\theta}\right)\left(\theta_h-\overline{\theta}\right)}{\omega\overline{\theta}(1-\pi)} - \frac{\overline{\theta}^2(1-\delta_0)}{2\psi}\right]\left[1+\frac{\overline{\theta}^2}{2}\frac{\omega(1-\pi)}{\psi\left(1-\overline{\theta}\right)}\right]
\end{aligned}
$$

Substitute i_A as per (48) and ϕ^* as per (49) then rearrange:

$$
\begin{aligned}
& \frac{\psi\left(1-\overline{\theta}\right)}{\omega(1-\pi)}\left[\theta_h - \frac{\overline{\theta}(1+\delta_0)}{2} + \left[\overline{\theta} + \frac{1-\theta_h}{\delta_0} + \frac{2}{\overline{\theta}}\frac{\psi\left(1-\overline{\theta}\right)}{\omega(1-\pi)}\right]\left[1-\frac{2\delta_0\left(\theta_h-\overline{\theta}\right)}{\overline{\theta}}\right]\right] \qquad (56) \\
< \; & \underbrace{\frac{\overline{\theta}}{\delta_0}\left[\frac{\overline{\theta}^2\delta_0^2}{4} + \frac{\delta_0}{2}\left[3\left(\theta_h-\overline{\theta}\right)(1-\theta_h)-\overline{\theta}^2\right] - \overline{\theta}(1-\theta_h)\right]}_{\text{call this } Z_2}
\end{aligned}
$$

Condition (56) will be true for $\frac{\psi}{\omega}$ sufficiently low if $Z_2 > 0$. Use $\overline{\theta} \equiv \pi\theta_\ell + (1-\pi)\theta_h$ to rewrite $Z_2 > 0$ as:

$$
\pi^2(\theta_h-\theta_\ell)^2 - 2\left[\theta_h + \frac{(2+3\delta_0)(1-\theta_h)}{\delta_0(2-\delta_0)}\right]\pi(\theta_h-\theta_\ell) + \theta_h\left[\theta_h + \frac{4(1-\theta_h)}{\delta_0(2-\delta_0)}\right] < 0
$$

Based on the roots of this quadratic, we can conclude that $Z_2 > 0$ requires:

$$
\pi(\theta_h-\theta_\ell) > \theta_h + \frac{(2+3\delta_0)(1-\theta_h)}{\delta_0(2-\delta_0)} - \sqrt{\frac{1-\theta_h}{2-\delta_0}\left(6\theta_h + \frac{(2+3\delta_0)^2(1-\theta_h)}{\delta_0^2(2-\delta_0)}\right)} \qquad (57)
$$

Condition (57) is satisfied by $\theta_\ell = 0$ and $\pi = 1$. The left-hand side is decreasing in θ_ℓ and increasing in π so it follows that $Z_2 > 0$ requires θ_ℓ sufficiently low and π sufficiently high.

The fourth step is to verify $W_j \leq X_j$. Use $W_j = \omega\xi_j$ and (23) with $\delta_1 = \omega$ to rewrite $W_j \leq X_j$ as:

$$
\xi_k \leq \frac{1-\delta_0}{\omega}
$$

Now use (47) with i_L^h as per (46) to replace ξ_k. Substitute i_A as per (48) and ϕ^* as per (49).

Rearrange to isolate all terms with $\frac{\psi\left(1-\overline{\theta}\right)}{\omega(1-\pi)}$ on one side. The condition for $W_j \leq X_j$ becomes:

$$\begin{aligned} & \frac{\psi\left(1-\overline{\theta}\right)}{\omega\left(1-\pi\right)}\left[\frac{\overline{\theta}^2}{2}+\left(1-\delta_0\right)\left[\overline{\theta}^2+\frac{\overline{\theta}\left(1-\theta_h\right)}{\delta_0}+2\frac{\psi\left(1-\overline{\theta}\right)}{\omega\left(1-\pi\right)}\right]\right] \\ \geq\ & \frac{\overline{\theta}^3}{4}\underbrace{\left[\left(1-\theta_h\right)\left(3-\frac{2}{\delta_0}\right)-\overline{\theta}\left(1-\frac{\delta_0}{2}\right)\right]}_{\text{call this } Z_3} \end{aligned} \tag{58}$$

If $Z_3 < 0$, then we can get $W_j \leq X_j$ without requiring a floor on $\frac{\psi}{\omega}$. This is useful since our other steps required $\frac{\psi}{\omega}$ sufficiently low. A sufficient condition for $Z_3 < 0$ is $\delta_0 \leq \frac{2}{3}$.

The fifth step is to verify $W_k \leq X_k$. Use $W_k = \omega\xi_k$ and (54) to rewrite $W_k \leq X_k$ as:

$$\xi_j \leq \frac{\delta_0}{\omega}$$

Now use (26) with $\alpha = \overline{\theta}$ and i_L^h as per (46) to replace ξ_j. Substitute i_A as per (48) and ϕ^* as per (49). Rearrange to isolate all terms with $\frac{\psi\left(1-\overline{\theta}\right)}{\omega(1-\pi)}$ on one side. The condition for $W_k \leq X_k$ becomes:

$$\begin{aligned} & \frac{\psi\left(1-\overline{\theta}\right)}{\omega\left(1-\pi\right)}\left[1-3\delta_0-\frac{2\delta_0}{\overline{\theta}}\left[\frac{1-\theta_h}{\delta_0}+\frac{2}{\overline{\theta}}\frac{\psi\left(1-\overline{\theta}\right)}{\omega\left(1-\pi\right)}\right]\right] \\ \leq\ & \frac{\overline{\theta}}{2}\underbrace{\left[\left(1-\theta_h\right)\left(3-\frac{1}{\delta_0}\right)-\overline{\theta}\left(\frac{1}{2}-\delta_0\right)\right]}_{\text{call this } Z_4} \end{aligned} \tag{59}$$

Condition (59) will be true for $\frac{\psi}{\omega}$ sufficiently low if $Z_4 > 0$. Use the definition of $\overline{\theta}$ to rewrite $Z_4 > 0$ as:

$$\pi\left(\theta_h-\theta_\ell\right)\delta_0\left(1-2\delta_0\right) > \theta_h\delta_0\left(1-2\delta_0\right)-2\left(1-\theta_h\right)\left(3\delta_0-1\right) \tag{60}$$

If $\delta_0 \geq \frac{1}{2}$, then (60) is always true. If $\delta_0 < \frac{1}{2}$, then (60) reduces to:

$$\theta_\ell < \frac{1}{\pi}\left[\frac{2\left(1-\theta_h\right)\left(3\delta_0-1\right)}{\delta_0\left(1-2\delta_0\right)}-\theta_h\left(1-\pi\right)\right]$$

This is a positive ceiling on θ_ℓ provided $\pi > 1-\frac{2(1-\theta_h)(3\delta_0-1)}{\theta_h\delta_0(1-2\delta_0)}$ with $\delta_0 > \frac{1}{3}$. Therefore, (60) is guaranteed by θ_ℓ sufficiently low, π sufficiently high, and $\delta_0 > \frac{1}{3}$.

The sixth step is to verify feasibility of $i_L^\ell = 0$. This requires $R_j + R_k \geq \overline{\theta}X_j + \theta_\ell X_k$. Use

(22) with $\alpha = \overline{\theta}$ to replace R_j. The desired inequality becomes:

$$R_k \geq \theta_\ell X_k + \omega \overline{\theta} \xi_j$$

Substituting X_k and R_k as per equations (54) and (55):

$$i_L^h \leq \frac{\theta_h - \theta_\ell}{\psi} \left[\delta_0 + \omega \left(\xi_k - \xi_j\right)\right]$$

Use (47) to replace ξ_k. Also use (26) with $\alpha = \overline{\theta}$ to replace ξ_j. Rearrange to isolate i_L^h then use (46) to replace i_L^h. Substitute i_A as per (48) and ϕ^* as per (49). Rearrange to isolate all terms with $\frac{\psi(1-\overline{\theta})}{\omega(1-\pi)}$ on one side. The feasibility condition for $i_L^\ell = 0$ becomes:

$$\frac{\psi\left(1-\overline{\theta}\right)}{\omega\left(1-\pi\right)} \left[\begin{array}{c} \frac{\overline{\theta}(5-\delta_0)}{4} - \left(\theta_h - \theta_\ell\right)\left[\frac{1-\theta_h}{\overline{\theta}} + \frac{2\delta_0 - 1}{2}\right] \\ + \frac{1-\theta_h}{2\delta_0} + \left[1 - \frac{2(\theta_h - \theta_\ell)\delta_0}{\overline{\theta}}\right] \frac{1}{\overline{\theta}} \frac{\psi\left(1-\overline{\theta}\right)}{\omega(1-\pi)} \end{array}\right] \leq \frac{3\overline{\theta}}{4} \underbrace{\left[\left(1-\theta_h\right)\left[\theta_h - \theta_\ell - \frac{\overline{\theta}}{\delta_0}\right] - \frac{\overline{\theta}^2}{2}\right]}_{\text{call this } Z_5} \quad (61)$$

Condition (61) will be true for $\frac{\psi}{\omega}$ sufficiently low if $Z_5 > 0$. Use the definition of $\overline{\theta}$ to rewrite $Z_5 > 0$ as:

$$\pi^2 \left(\theta_h - \theta_\ell\right)^2 - 2\left[\pi\theta_h + \frac{\left(\pi + \delta_0\right)\left(1-\theta_h\right)}{\delta_0}\right]\left(\theta_h - \theta_\ell\right) + \theta_h\left[\theta_h + \frac{2\left(1-\theta_h\right)}{\delta_0}\right] < 0$$

Based on the roots of this quadratic, we can conclude that $Z_5 > 0$ requires:

$$\theta_\ell < \frac{1}{\pi^2}\left[\sqrt{2\pi\theta_h\left(1-\theta_h\right) + \frac{\left(\pi+\delta_0\right)^2\left(1-\theta_h\right)^2}{\delta_0^2}} - \frac{\left(\pi+\delta_0\right)\left(1-\theta_h\right)}{\delta_0} - \theta_h\pi\left(1-\pi\right)\right]$$

This is a positive upperbound on θ_ℓ provided $\frac{\theta_h(1-\pi)^2}{2(1-\theta_h)} + \frac{1-\pi}{\delta_0} < 1$. Therefore, $Z_5 > 0$ requires θ_ℓ sufficiently low and π sufficiently high.

It now remains to check that the interbank rate increases when moving from $\alpha = 0$ to $\alpha = \overline{\theta}$. This requires (46) to exceed (30) or, equivalently:

$$\frac{\left(1+i_A\right)^2 - 1}{1-\pi} > \left(1-\delta_0\right)\left[\frac{\overline{\theta}}{\psi} + \frac{4\left(1-\overline{\theta}\right)}{\omega\overline{\theta}\left(1-\pi\right)} \frac{2\left(1-\overline{\theta}\right) + \phi^*\omega}{2\left(1-\overline{\theta}\right) + 3\phi^*\omega}\right]$$

Using i_A as per (48) and ϕ^* as per (49):

$$\frac{\psi\left(1-\overline{\theta}\right)}{\omega\left(1-\pi\right)}\left[\frac{1-\theta_h}{\delta_0}+\frac{\overline{\theta}\left(1-2\delta_0\right)}{2\left(1-\delta_0\right)}+\frac{2}{\overline{\theta}}\frac{\psi\left(1-\overline{\theta}\right)}{\omega\left(1-\pi\right)}\right]<\frac{3\overline{\theta}^2}{4\left(1-\delta_0\right)}\left[1-\theta_h+\frac{\overline{\theta}\delta_0}{2}\right] \tag{62}$$

The right-hand side is positive so (62) will be true for $\frac{\psi}{\omega}$ sufficiently low.

Putting everything together, we have shown that the model with endogenous funding shares generates the desired results under the following conditions: π sufficiently high, θ_ℓ and $\frac{\psi}{\omega}$ sufficiently low, $\delta_0 \in \left(\frac{1}{3}, \frac{2}{3}\right)$, and i_A as per (48). The results then extend to a non-empty range of i_A by continuity. □

Comparison We now compare the interbank rate increases in the fixed share and endogenous share models. Notice from the proof of Proposition 4 that the interbank rate at $\alpha = 0$ is the same in both models. Therefore, we just need to show that the interbank rate in the endogenous share model exceeds the interbank rate in the fixed share model at $\alpha = \overline{\theta}$. In other words, we need to show that (46) exceeds (39) for a given set of parameters. This reduces to:

$$\frac{\left(1+i_A\right)^2-1}{1-\pi}\left[1-\frac{\phi^*}{\frac{2\left(1-\overline{\theta}\right)}{\omega}+\frac{\overline{\theta}^2\left(1-\pi\right)}{\psi}}\right]<\frac{\overline{\theta}\left(1-\delta_0\right)}{\psi}$$

which is exactly (52), where (52) was the condition for $\xi_k > 0$ at $\alpha = \overline{\theta}$ in the endogenous share model. To complete the proof, we must now show that there are indeed parameters that satisfy the conditions in both models. For $\alpha = 0$, we imposed conditions (32) and (35) along with π sufficiently high and θ_ℓ sufficiently low. These conditions applied to both models. For $\alpha = \overline{\theta}$ in the fixed share model, we also imposed conditions (40), (41), (42), and (43) along with $\frac{\psi}{\omega}$ sufficiently low. For $\alpha = \overline{\theta}$ in the endogenous share model, we added $\delta_0 \in \left(\frac{1}{3}, \frac{2}{3}\right)$ and i_A in the neighborhood of (48). In (50) and (51), we showed that π sufficiently high and θ_ℓ sufficiently low make (48) satisfy condition (35). We have also shown that condition (43) is stricter than condition (32). Therefore, we just need to show that (48) satisfies conditions (40), (41), (42), and (43). Substituting i_A as per (48) into these conditions produces the following inequalities which we must check:

$$\frac{\psi\left(1-\overline{\theta}\right)}{\omega\left(1-\pi\right)}>\frac{\overline{\theta}^2\left(2\delta_0-1\right)}{4\left(1-\delta_0\right)} \tag{63}$$

$$\theta_\ell<\left[1-\frac{1}{\pi\left(1+2\delta_0\right)}\right]\theta_h \tag{64}$$

$$\frac{\psi\left(1-\overline{\theta}\right)}{\omega\left(1-\pi\right)}\left[1-\frac{2\left(\theta_h-\theta_\ell\right)\delta_0}{\overline{\theta}}\right]<\overline{\theta}^2\left[\frac{\left(\theta_h-\theta_\ell\right)\delta_0}{\overline{\theta}}-1\right] \tag{65}$$

$$\frac{\psi\left(1-\overline{\theta}\right)}{\omega\left(1-\pi\right)}<\frac{\overline{\theta}^2\delta_0}{2\left(1-\delta_0\right)} \tag{66}$$

A sufficient condition for (63) is $\delta_0 \leq \frac{1}{2}$ which is still consistent with $\delta_0 \in \left(\frac{1}{3}, \frac{2}{3}\right)$. Condition (64) is just another positive upperbound on θ_ℓ provided $\pi > \frac{1}{1+2\delta_0}$. In other words, (64) is satisfied by θ_ℓ sufficiently low and π sufficiently high. Condition (65) will be true for $\frac{\psi}{\omega}$ sufficiently low if $\left(\theta_h-\theta_\ell\right)\delta_0 > \overline{\theta}$ or, equivalently, $\theta_\ell < \left[1-\frac{1}{\delta_0+\pi}\right]\theta_h$ with $\pi > 1-\delta_0$ which again means θ_ℓ sufficiently low and π sufficiently high. Finally, condition (66) is clearly satisfied by $\frac{\psi}{\omega}$ sufficiently low. □ ■

Proof of Proposition 6

Evaluate (26) at $\alpha = \overline{\theta}$ then subtract (47) to get:

$$\xi_j - \xi_k \overset{sign}{=} \left[\frac{\left(1+i_A\right)^2-1}{1-\pi}-\frac{\overline{\theta}\left(1-\delta_0\right)}{\psi}\right]+2\left[\frac{\left(1+i_A\right)^2-1}{1-\pi}-i_L^h\right]$$

The expression in the first set of square brackets is positive by condition (32). The expression in the second set of square brackets is proportional to μ_j. The proof of Proposition 5 established $\mu_j > 0$. Therefore, $\xi_j > \xi_k$ at $\alpha = \overline{\theta}$.

Now consider total credit:

$$TC \equiv 1 - R_j - R_k$$

Use market clearing as per (12) to replace $R_j + R_k$:

$$TC = 1 - \overline{\theta}X_j - \theta_h X_k + \psi i_L^h$$

Use (23) and (24) to replace X_j and X_k:

$$TC = 1 - \overline{\theta} - \left(\theta_h - \overline{\theta}\right)\delta_0 + \delta_1\left(\theta_h - \overline{\theta}\right)\left(\xi_j - \xi_k\right) + \psi i_L^h$$

Proposition 5 showed $i_L^h\big|_{\alpha=\overline{\theta}} > i_L^h\big|_{\alpha=0}$. We also know $\xi_j = \xi_k = 0$ at $\alpha = 0$ and $\xi_j > \xi_k$ at $\alpha = \overline{\theta}$. Therefore, we can conclude $TC|_{\alpha=\overline{\theta}} > TC|_{\alpha=0}$.

Finally, we want to show that the loan-to-deposit ratios of big and small banks converge. The equilibrium has $\tau_j = 1$, meaning that small banks move all WMPs (and the associated investments) off-balance-sheet. The loan-to-deposit ratio of the representative small bank is then $\lambda_j \equiv 1 - \frac{R_j}{X_j - W_j}$. The equilibrium also has $\tau_k = 0$, meaning that the big bank records everything on-balance-sheet. Its loan-to-deposit ratio is then $\lambda_k \equiv 1 - \frac{R_k}{X_k}$. Proposition 4 established $R_k > 0 = R_j$ at $\alpha = 0$ so it follows that $\lambda_k|_{\alpha=0} < 1 = \lambda_j|_{\alpha=0}$. To show convergence, we just need to show $\lambda_k|_{\alpha=\overline{\theta}} > \lambda_k|_{\alpha=0}$ since $\lambda_j|_{\alpha=\overline{\theta}} < \lambda_j|_{\alpha=0}$ follows immediately from equation (22). Use $X_j + X_k = 1$ along with the definition of λ_k to rewrite (12) as:

$$\psi i_L^h = \overline{\theta} + \left[\theta_h - \overline{\theta} - (1 - \lambda_k)\right] X_k - R_j$$

We know $i_L^h\big|_{\alpha=\overline{\theta}} > i_L^h\big|_{\alpha=0}$ so it must be the case that:

$$\left[\theta_h - \overline{\theta} - (1 - \lambda_k|_{\alpha=\overline{\theta}})\right] X_k|_{\alpha=\overline{\theta}} - R_j|_{\alpha=\overline{\theta}} > \left[\theta_h - \overline{\theta} - (1 - \lambda_k|_{\alpha=0})\right] X_k|_{\alpha=0}$$

Proposition 4 also established $\xi_j = \xi_k = 0$ at $\alpha = 0$. Substituting into equation (24) then implies $X_k = \delta_0$ at $\alpha = 0$ so:

$$\lambda_k|_{\alpha=\overline{\theta}} \frac{X_k|_{\alpha=\overline{\theta}}}{\delta_0} - \lambda_k|_{\alpha=0} > \underbrace{\frac{R_j|_{\alpha=\overline{\theta}}}{\delta_0} - \left[1 - \pi\left(\theta_h - \theta_\ell\right)\right]\left[1 - \frac{X_k|_{\alpha=\overline{\theta}}}{\delta_0}\right]}_{\text{call this } Z_6}$$

We have shown $\xi_j > \xi_k$ at $\alpha = \overline{\theta}$ so equation (24) also implies $\frac{X_k|_{\alpha=\overline{\theta}}}{\delta_0} \leq 1$ for any $\delta_1 \geq 0$. Therefore, $Z_6 \geq 0$ will be sufficient for $\lambda_k|_{\alpha=\overline{\theta}} > \lambda_k|_{\alpha=0}$. If $\delta_1 = 0$, then $Z_6 \propto R_j|_{\alpha=\overline{\theta}} \geq 0$. If $\delta_1 = \omega$, then we can rewrite $Z_6 \geq 0$ as:

$$1 - \delta_0 - \omega\xi_k \geq \frac{1 - \pi\left(\theta_h - \theta_\ell\right)}{\overline{\theta}} \omega\left(\xi_j - \xi_k\right) \tag{67}$$

where ξ_j is given by (26) with $\alpha = \overline{\theta}$ and ξ_k is given by (47). Use these expressions to substitute out ξ_j and ξ_k then use equation (46) to substitute out i_L^h. Evaluate i_A at (48) and ϕ^* at (49) to rewrite (67) as:

$$\begin{aligned}
&\frac{4\psi\left(1 - \overline{\theta}\right)\left(1 - \delta_0\right)}{\omega\overline{\theta}\left(1 - \pi\right)} + \underbrace{\overline{\theta}\left(2 - 3\delta_0\right) + \left(1 - \theta_h\right)\left(\frac{2}{\delta_0} - 3 - \delta_0\right)}_{\text{call this } \Delta(\delta_0)} \\
\geq \quad & -\frac{\overline{\theta}^2}{4}\frac{\omega\left(1 - \pi\right)}{\psi\left(1 - \overline{\theta}\right)}\underbrace{\left[2\overline{\theta}\left(1 - 2\delta_0\right) + \left(1 - \theta_h\right)\left(\frac{4}{\delta_0} - 6 - 3\delta_0\right)\right]}_{\text{call this } \widetilde{\Delta}(\delta_0)}
\end{aligned}$$

A sufficient condition for this is $\min\left\{\Delta\left(\delta_0\right), \widetilde{\Delta}\left(\delta_0\right)\right\} \geq 0$. Notice $\Delta'\left(\cdot\right) < 0$ and $\widetilde{\Delta}'\left(\cdot\right) < 0$. Also notice $\min\left\{\Delta\left(\frac{1}{2}\right), \widetilde{\Delta}\left(\frac{1}{2}\right)\right\} > 0$ and $\min\left\{\Delta\left(\frac{2}{3}\right), \widetilde{\Delta}\left(\frac{2}{3}\right)\right\} < 0$. Therefore, there is a threshold $\overline{\delta_0} \in \left(\frac{1}{2}, \frac{2}{3}\right)$ such that $\delta_0 \leq \overline{\delta_0}$ guarantees $Z_6 \geq 0$. ■

Appendix B – Deposit and WMP Demands

Here we sketch a simple household maximization problem which generates the demands in equations (1) and (2). There is a continuum of ex ante identical households indexed by $i \in [0, 1]$. Each household is endowed with X units of funding. Let D_{ij} and W_{ij} denote the deposits and WMPs purchased by household i from bank j, where:

$$\sum_j \left(D_{ij} + W_{ij}\right) \leq X \tag{68}$$

Assume that buying W_{ij} entails a transaction cost of $\frac{1}{2\omega_0} W_{ij}^2$, where $\omega_0 > 0$. As per the main text, the rate of return on the WMP is zero if withdrawn early and ξ_j otherwise. The rate of return on deposits is always zero and the average probability of early withdrawal is $\overline{\theta}$. The household requires subsistence consumption of X in each state, above which it is risk neutral. If the household were to bypass the banking system and invest in long-term projects directly, it would fall below subsistence in the state where it needs to liquidate early. Therefore, the household does not invest directly. Instead, it chooses D_{ij} and W_{ij} for each j to maximize:

$$\sum_j \left(D_{ij} + \left[1 + \left(1 - \overline{\theta}\right)\xi_j\right] W_{ij} - \frac{W_{ij}^2}{2\omega_0}\right)$$

subject to (68) holding with equality.[40] The first order condition with respect to W_{ij} is:

$$W_{ij} = \left(1 - \overline{\theta}\right)\omega_0 \xi_j \tag{69}$$

Substituting (69) into (68) when the latter holds with equality gives the household's total deposit demand, $D_i \equiv \sum_j D_{ij}$. The household is indifferent about the allocation of D_i across

[40] Here is how to recover the two-point distribution of idiosyncratic bank shocks in Section 4.1 from the household withdrawals. Each household has probability θ_ℓ of being hit by an idiosyncratic consumption shock at $t = 1$ and having to withdraw all of its funding early. This results in each bank losing fraction θ_ℓ of its deposits and WMPs at $t = 1$. Then $\theta_h - \theta_\ell$ of the remaining $1 - \theta_\ell$ households observe a sunspot and withdraw all of their funding from $1 - \pi$ banks at $t = 1$. The $\theta_h - \theta_\ell$ households and $1 - \pi$ banks involved in the sunspot are chosen at random. Note $\overline{\theta} \equiv \pi\theta_\ell + (1 - \pi)\theta_h$.

banks so we assume that it simply allocates D_i uniformly. For J banks, this yields:

$$D_{ij} = \frac{X}{J} - \frac{\left(1-\overline{\theta}\right)\omega_0}{J}\xi_j - \frac{(J-1)\left(1-\overline{\theta}\right)\omega_0}{J}\frac{1}{J-1}\sum_{x\neq j}\xi_x \tag{70}$$

With a unit mass of ex ante identical households, $W_j = W_{ij}$ and $D_j = D_{ij}$. As J approaches a unit mass of equally-weighted banks, (69) and (70) belong to the family of functions specified by (1) and (2).

Appendix C – Benchmark with Aggregate Shock

Consider the benchmark model (only price-taking banks) in Sections 4.1 and 4.2 but with an aggregate interbank shock. In particular, the interbank rate is i_L^ℓ with probability π and i_L^h with probability $1-\pi$. The expected interbank rate is $i_L^e \equiv \pi i_L^\ell + (1-\pi)\, i_L^h$. We will specify how i_L^ℓ and i_L^h are determined shortly. In the meantime, banks take both as given.

The objective function of the representative bank simplifies to:

$$\Upsilon_j = (1+i_A)^2 (X_j - R_j) + (1+i_L^e) R_j - \left[X_j + i_L^e\overline{\theta}X_j + \left(1-\overline{\theta}\right)\xi_j W_j\right] - \tfrac{\phi}{2}X_j^2$$

This is identical to the benchmark model except with the expected interbank rate i_L^e instead of the deterministic i_L. Therefore, the first order conditions are still given by equations (7) to (9) but with i_L^e in place of i_L.

The goal is to show that i_L^e is always highest at $\alpha = 0$. The proof follows Proposition 3 but, to proceed, we must replace the deterministic market clearing condition (equation (4)) with conditions for each realization of the aggregate shock. We model the shock as a shock to the aggregate demand for liquidity at $t = 1$. In particular, aggregate liquidity demand is $\overline{\theta}X - \varepsilon$ with probability π and $\overline{\theta}X$ with probability $1-\pi$, where $\varepsilon > 0$. The interbank rates are then i_L^ℓ and i_L^h respectively. To avoid liquidity shortages, we need these rates to satisfy:

$$R_j + \psi i_L^\ell \geq \overline{\theta}X - \varepsilon \tag{71}$$

$$R_j + \psi i_L^h \geq \overline{\theta}X \tag{72}$$

The equilibrium i_L^h solves (72) with equality. If $i_L^h \leq \frac{\varepsilon}{\psi}$, then we can set $i_L^\ell = 0$. Otherwise, the equilibrium i_L^ℓ solves (71) with equality.

Let i^e_{L0} denote the expected interbank rate at $\alpha = 0$ and let $i^e_{L1}(\rho_2)$ denote the expected interbank rate at some $\alpha > 0$. Using (71) and (72), we can write:

$$i^e_{L1}(\rho_2) = \frac{\overline{\theta} X}{\psi} - \frac{R_{j1}(\rho_2)}{\psi} - \frac{\pi}{\psi} \min\left\{\overline{\theta} X - R_{j1}(\rho_2), \varepsilon\right\} \tag{73}$$

where $R_{j1}(\rho_2)$ is reserve holdings at the $\alpha > 0$ being considered. The proof of $i^e_{L1}(\rho_2) \leq i^e_{L0}$ proceeds by contradiction. In particular, suppose $i^e_{L1}(\rho_2) > i^e_{L0}$. Then (7) implies $\mu_j > 0$ at $\alpha = 0$. Complementary slackness then implies $R_j = 0$ at $\alpha = 0$ so we can write:

$$i^e_L = \frac{\overline{\theta} X}{\psi} - \frac{\pi}{\psi} \min\left\{\overline{\theta} X, \varepsilon\right\} \tag{74}$$

Subtract (74) from (73) to get:

$$i^e_{L1}(\rho_2) = i^e_{L0} - \frac{R_{j1}(\rho_2)}{\psi} + \frac{\pi}{\psi}\left[\min\left\{\overline{\theta} X, \varepsilon\right\} - \min\left\{\overline{\theta} X - R_{j1}(\rho_2), \varepsilon\right\}\right]$$

There are three cases. If $\varepsilon \leq \overline{\theta} X - R_{j1}(\rho_2)$, then:

$$i^e_{L1}(\rho_2) = i^e_{L0} - \frac{R_{j1}(\rho_2)}{\psi}$$

If $\overline{\theta} X - R_{j1}(\rho_2) < \varepsilon < \overline{\theta} X$, then:

$$i^e_{L1}(\rho_2) = i^e_{L0} - \frac{1-\pi}{\psi} R_{j1}(\rho_2) - \frac{\pi}{\psi}\left(\overline{\theta} X - \varepsilon\right)$$

If $\overline{\theta} X \leq \varepsilon$, then:

$$i^e_{L1}(\rho_2) = i^e_{L0} - \frac{1-\pi}{\psi} R_{j1}(\rho_2)$$

In each case, $i^e_{L1}(\rho_2) > i^e_{L0}$ would require $R_{j1}(\rho_2) < 0$ which is impossible. ■