The Rise of China's Shadow Banking System*

Kinda Hachem
Chicago Booth and NBER

Zheng Michael Song
Chicago Booth

First Draft: January 2015
This Draft: May 2015

Abstract

Shadow banking in China has grown very rapidly during the past decade. This paper studies the causes and impending consequences. At a general level, we provide the first jointly empirical and theoretical analysis of shadow banking in China. At a more granular level, we uncover a novel link between shadow banking and interbank market power which connects an otherwise disperse set of facts.

*We thank Anil Kashyap for several helpful conversations. We also thank Doug Diamond, Chang-Tai Hsieh, John Huizinga, Mark Kruger, 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, and NUS. Both authors gratefully acknowledge financial support from Chicago Booth.

1 Introduction

Regulatory arbitrage is transforming China's financial system. The 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. Second, it restricts how many loans can be made and to whom. A new form of intermediation – shadow banking – has recently emerged in China to circumvent these regulations. Lending by China's shadow sector has grown very rapidly yet the causes and potential consequences of this growth remain poorly understood.[1] Our paper fills the gap in three steps. We first document the institutional features that have contributed to the rise of shadow banking in China. We then present a new and comprehensive set of empirical facts that emerged alongside China's shadow sector. Finally, we develop a model that connects the dots between these facts. Our analysis reveals a novel interaction between big and small banks that is making government regulation particularly counterproductive.

To better understand shadow banking in China, it is necessary to understand the Chinese system more broadly. Over the past 25 years, a fast-growing private sector has emerged.[2] Non-state-owned banks were keen to serve the private sector but were kept small by limited retail networks and an uncompetitive deposit rate ceiling imposed by the government. Until recently, the interbank market provided a solution: these small banks channeled almost all of their existing deposits into non-financial loans then borrowed from big state-owned banks when in need of extra/emergency liquidity. The big banks were flush with deposits and willing to lend to small banks at an appropriate rate rather than make additional loans to

[1] Our conservative estimate is that China's shadow banking system grew from a negligible fraction of GDP in 2007 to 16% of GDP in 2014. A less conservative estimate comes in at 5% of GDP in 2007 and 24% of GDP in 2014. Section 2.4 explains how we construct these estimates. See Elliott et al (2015) for a survey of other estimates. While there is a distribution of estimates, all point to very rapid growth.

[2] See, for example, Hsieh and Klenow (2009), Song et al (2011), Brandt et al (2012), Lardy (2014), Hsieh and Song (2015), and the references therein.

the tepid state sector or the politically thorny private sector.

Things changed in the late 2000s 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. We exploit cross-sectional differences to show that China's shadow banking sector emerged to circumvent enforcement of the 75% cap. In particular, small banks responded by offering a new savings instrument called a "wealth management product" or WMP for short. 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.[3]

WMPs have two additional features that make them potentially problematic. First, there is a mismatch in the funding arrangements at the trust companies. 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. WMPs thus bear an unsettling resemblance to the asset-backed commercial paper vehicles that collapsed during the 2007-2009 financial crisis.[4]

Second, WMPs are not subject to interest rate restrictions. This introduces a tension with China's big banks. By broaching the uncompetitive ceiling on deposit rates, WMPs have the potential to poach a lot of savings from big banks. The big banks can prevent this by offering their own high-return WMPs. However, since big banks are not looking to

[3]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).

[4]For more on the 2007-2009 crisis, see Brunnermeier (2009), Gorton and Metrick (2012), Covitz et al (2013), Kacperczyk and Schnabl (2013), Krishnamurthy et al (2014), and the references therein.

arbitrage the 75% cap, they prefer cheap deposits over more expensive WMPs. We argue that this has led big banks to change their interbank behavior. Given the fragile nature of WMPs, the interbank market remains an important source of emergency liquidity for small banks. Therefore, by signalling that they will not provide such liquidity, big banks can make small banks less aggressive in WMP issuance and lessen the extent to which small banks poach savings. This strategy by the big banks is making China's interbank markets tighter and more volatile. In the interbank repo market, for example, interest rates increased from an average of 2.8% in 2008 to an average of 4.4% in 2013, reaching an unprecedented 11.6% in mid-2013. We dig deeper into the mid-2013 event and show that big banks 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 is taking place in China. First, the reallocation of savings from deposits at the big banks to higher-return WMPs at the small banks increases 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.

The rest of the paper proceeds as follows. Section 2 describes the basic features of China's banking system, Section 3 presents our empirical evidence, Section 4 builds a model to rationalize the facts, and Section 5 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.

First, the state-owned banks became much more profit-driven. All four of these banks went through a major restructuring in the mid-2000s and are now publicly listed. Their average non-performing loan ratio has not exceeded 2% since 2008, down significantly from 30% in 2000. Combined profits have also been remarkable, growing 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.[5] To be sure, the government still maintains close ties with the Big Four: the Ministry of Finance and Huijin (a government-owned investment company) retain controlling interests while the Organization Department of the Chinese Communist Party directly appoints top executives.[6] These ties limit how intensely the Big Four compete against each other but, beyond that, the government no longer interferes with day-to-day banking operations.

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

[6] Similar features are found in big state-owned firms in the industrial sector (Hsieh and Song (2015)).

The second notable change that followed China's market-oriented reforms was entry of small and medium-sized commercial banks. China now has ten joint-stock commercial banks operating nationally and over one hundred city banks specializing in regional business. These banks are individually small when compared to the Big Four. For example, deposits at the average joint-stock bank were barely 17% of deposits at the average Big Four in 2013. However, as a group, 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 deposits so even the major restructuring of the mid-2000s did not stabilize their market share.[7]

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. Together, CBRC and the central bank oversee a regulatory environment with three main pillars. One pillar, capital regulation, has been shaped by the international Basel Accords.[8] The other two pillars – a ceiling on bank deposit rates and a cap on bank loan-to-deposit ratios – are not based on similar international standards.

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.

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

[8] 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%.

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.

The next important regulation, a 75% loan-to-deposit cap, is formally written into China's Law on Commercial Banks. The law was enacted in 1995 but enforcement of the 75% cap was initially loose. Things changed around 2008 when CBRC began stricter enforcement to rein in the fast-growing loans made by small and medium-sized banks. Enforcement became more formal in 2011 when CBRC established China's "CARPALs" regulatory system. The loan-to-deposit ratio is one of thirteen supervised indicators, with CBRC now monitoring the average daily ratio rather than just the end-of-year or end-of-month ratio. Stricter enforcement was also complemented by a rapid increase in the reserve requirements set by the central bank. Official requirements went from 9% in early 2007 to 15.5% in February 2010. They were then increased twelve times to reach 21.5% by December 2011.

2.3 Bank-Trust Cooperation as Regulatory Arbitrage

With the main banking regulations in hand, let us now address how these regulations are being circumvented. Wealth management products (WMPs) are the centerpiece of regulatory arbitrage in China. WMPs outstanding ballooned from 2% of GDP in 2007 to nearly 25% of GDP in 2014. A WMP is a savings instrument that is typically sold at bank counters. These products have two features which help them get around the regulations discussed above. First, WMP returns are not subject to the 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.[9] Second, WMPs do not have to be principal-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

[9] Almost all WMPs delivered yields above or equal to their promised yields.

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, 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. In March 2013, CBRC went even further and announced that WMPs could invest at most 35% in non-standard debt assets. Most trust activity is considered non-standard debt so the response was a second wave of shadow banking which operates as per Figure 5. In short, trust companies offer up beneficiary rights which make their way to banks via "offline" repos.[10] WMPs deriving their returns from these rights are then advertised as WMPs backed by interest rate products, not WMPs backed by trust assets.

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 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.[11] Further support for the

[10] Offline transactions are ones which do not go through the China Foreign Exchange Trade System.

[11] The sectorial composition of trust company assets has become more even over time, with infrastructure

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.[12]

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. Recall that 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, we use the widely-cited data on total social financing constructed by China's National Bureau of Statistics. 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.[13] 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

and real estate projects losing ground to industrial and commercial enterprises.

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

[13] 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.

entrusted loans into one measure of shadow banking.[14] By this measure, shadow banking 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 accounts for a sizeable amount of the broader measure considered here.

3 Empirical Evidence

This section establishes the core facts that motivate our paper. We first show that loan-to-deposit rules triggered shadow banking among China's small and medium-sized banks (henceforth SMBs). We then show that China's four biggest banks (the Big Four) have become more aggressive in traditional lending and are manipulating interbank markets. Section 4 will demonstrate that these facts are linked in a model of asymmetric competition. In the meantime, 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 loan-to-deposit ratio across all commercial banks averaged 67% between 2007 and 2013 so the 75% cap described in Section 2 does not seem binding at an aggregate level. The cross-section, however, reveals a different story. Figure 6 plots the evolution of raw loan-to-deposit ratios by bank size.[15] As a group, SMBs are constrained by the 75% cap and are reducing their loan-to-deposit ratios to comply. The ten national joint-stock banks, for example, had an average ratio of 74% between 2007 and 2013. The Big Four are not similarly

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

[15] Data for Figures 6 and 8 come from the central bank (PBOC) which restricts SMBs to nationally-operating banks with total assets below RMB 2 trillion in 2008.

constrained: their loan-to-deposit ratio has not exceeded 65% for at least a decade.[16] We exploit this cross-sectional difference in what follows.

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 bank-trust cooperation. We confirm these predictions here.

Between 2008 and 2014, the Big Four accounted for 27% of all new WMP batches. The most they ever accounted for in a given year was 31% while the least was 18%. Batch statistics 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 big banks accounted for around 40% of WMP balances outstanding at the end of both 2012 and 2013. This is in contrast to a market share of 50% when it comes to traditional deposits. Therefore, regardless of whether we use the count data or the balance data, SMBs are disproportionately more involved in WMP issuance.

SMBs are also disproportionately more involved in non-guaranteed WMPs. Between 2008 and 2014, the Big Four accounted for 23% of new non-guaranteed WMP batches. The most they ever accounted for in a given year was 37% while the least was 14%. Notice that the Big Four's 23% share of non-guaranteed batches is lower than their 27% share of all batches. The Big Four thus have a lower intensity of non-guaranteed issuance than the SMBs. To this point, we find that the Big Four issued 46% of their WMP batches without a guarantee while SMBs issued 57% of their WMP batches without a guarantee between 2008 and 2014.

[16] A common story is that the government uses individual loan quotas to impose even stricter limits on big banks. However, many of the banking insiders we spoke with conceded that quotas are open to negotiation and that the Big Four have enough sway to loosen any quotas imposed on them.

The gap widens to 62% for SMBs versus 43% for the Big Four in the second half of our sample. Aggregating across banks, roughly 56% of all WMP batches between 2008 and 2014 were issued without a guarantee. Figure 7 decomposes this percentage by issuer type and minimum purchase amount (i.e., summing all the bars in Figure 7 returns 56%). For any minimum purchase amount, SMBs issued more non-guaranteed batches than the Big Four.[17]

Figure 8 then shows a dramatic rise in "other investments" by SMBs as CBRC begins cracking down on bank-trust cooperation. Other investments include purchases of trust beneficiary rights or holdings of such rights through reverse repos. Figure 8 also shows that SMBs recorded a big jump in deposits by banks and other financial institutions, alternatively called placements from counterparts. Along with using trust beneficiary rights to generate returns for its own WMPs, a bank can use trust beneficiary rights to generate returns for other institutions (see Figure 5). In contrast, there is no rise in other investments or placements from counterparts at the Big Four.

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 (Figure 9). 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 a binding ceiling on deposit rates. This 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 new 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

[17] Data limitations prevent an exact decomposition by RMB value but we estimate that the Big Four accounted for *at most* 42% of non-guaranteed WMP balances outstanding at the end of 2013. We say at most because the entire WMP balance reported by Bank of China is described as an unconsolidated balance yet the micro data captures several guaranteed batches for this bank.

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.[18] We thus supplement the cross-sectional evidence with a case study of CMB. As will be explained below, the evolution of CMB's product characteristics provides further evidence that WMPs are a response to loan-to-deposit rules.

Figure 10 illustrates CMB's loan-to-deposit ratio. The blue line is the raw ratio measured at the end of the year. The solid gray line is the official ratio used by regulators. Relative to the raw ratio, the official measure excludes certain agricultural and micro loans. Prior to 2007, bills were also not part of CMB's official calculation. On the surface, CMB's official loan-to-deposit ratio was visibly below the 75% cap when stricter enforcement began in 2008. However, as the dashed gray line in Figure 10 reveals, the cap was binding once regulators drilled down to only RMB-denominated activities.

The subsequent growth in CMB's wealth management products was dramatic. As shown in Figure 11(a), 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

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

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.[19] The use of credit and notes as backing assets has since fallen due to CBRC's rules on bank-trust cooperation. CMB now backs most of its WMPs with interest rate products, engaging in the trust beneficiary right business discussed in Section 2 (see Figure 11(b)).

Further evidence on the importance of loan-to-deposit rules comes from changes in WMP maturity. Figure 12 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 decrease in the promised annualized yield on non-guaranteed products. Instead, the drop in CMB's non-guaranteed maturity lines up very well with changes in how CBRC monitored loan-to-deposit ratios. 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. Upon maturity, the principal and interest from a non-guaranteed (off balance sheet) WMP are automatically transferred to the buyer's deposit account. Reserves and deposits rise, lowering the loan-to-deposit ratio observed by CBRC.[20] In the first half of 2011, CMB's non-guaranteed products had a median maturity of just under 1 month which is good timing for end-of-month exams. However, once CBRC began monitoring average daily ratios in mid-2011, arbitraging on maturity became much harder. According to Figure 12, CMB's median non-guaranteed maturity has since returned to almost 3 months.

[19] 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%).

[20] Keeping the automatic deposits as reserves is only one approach. Another is to bring loans back on balance sheet through the repo market. The idea is similar to Figure 5 but with very short-term repos (i.e., only around exam dates). The data suggest that CMB just kept reserves between 2009 and 2011.

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 33 trillion of new loans between 2007 and 2013. Deposits increased by RMB 46 trillion over the same period, pushing the aggregate loan-to-deposit ratio from 64.7% in 2007 to 69.5% in 2013. Consolidating the traditional and shadow sectors, we estimate that the ratio of total credit to total savings increased by at least 10 percentage points between 2007 and 2013.

Interestingly, the Big Four are driving the rising aggregate loan-to-deposit ratio in the traditional sector. This is consistent with Figure 6 which shows that the Big Four's loan-to-deposit ratio was falling prior to 2008 but has been rising ever since. The recent rise reflects both higher loan growth and lower deposit growth. For the three years between 2006 and 2008, loans and deposits at the Big Four grew at average annual rates of 10.8% and 13.9% respectively. For the four years between 2011 and 2014, these rates were 11.8% and 8.8% respectively. The increase in loan growth was not driven by one particular year: the coefficient of variation for loan growth fell from 0.31 to 0.08 between the two periods while the coefficient of variation for deposit growth was around 0.3 in both periods.

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. We view this as an incomplete explanation. First, it does not account for high loan growth since 2011. Second, it accounts for at most a modest fraction of big bank lending in 2009 and 2010. Much of the stimulus package was to be borrowed by local governments and, 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. Nevertheless, gross loans at the Big Four jumped by RMB 4.8 trillion between 2008 and 2009 then by another RMB 3.5 trillion between 2009 and 2010. There is thus a sizable jump in big bank lending that cannot be explained by forced financing of the stimulus package.

We argue instead that the Big Four have become strategically less liquid to tighten interbank conditions and put pressure on the shadow banking activities of SMBs. Section 3.3 provides empirical evidence that big banks are indeed manipulating the interbank market against SMBs. Section 4 will then formalize our argument with a model.

3.3 Price Manipulation on Interbank Markets

To set the stage, Figure 13 shows an upward trend in monthly average interbank interest rates since 2009 despite generally less contractionary monetary policy 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.[21] 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.[22] Most of the seasonal pressures seemed to have subsided yet the market re-ignited 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

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

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

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 14(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 14(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 15, 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.[23] City and 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

[23] For completeness, the overnight and 7 day maturities shown in Figure 15 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.

the Big Four, overnight trades accounted for 96% of big bank borrowing but only 83% of big bank lending to non-policy banks. For comparison, 79.4% of policy bank lending to banks outside the Big Four was 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 16 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.[24] 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 re-ignited 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.[25]

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

[25] Now that we have seen the details, a simple way to summarize the importance of the Big Four is to report how the overnight repo rate correlates with lending by big banks versus policy banks. In June 2013, the correlation with lending by big banks to JSCBs was -0.38 while the correlation with lending by policy banks to JSCBs was 0.13. A similar pattern emerges for lending to other SMBs: the correlation between

4 Model

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 counterproductive.

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) maturity transformation, (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 and is novel relative to workhorse banking models. To isolate the contribution of this ingredient, 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. We then extend the model in Section 4.5 to discuss the recent regulatory crackdown on bank-trust cooperation.

4.1 Environment

There are three periods, $t \in \{0, 1, 2\}$, and a continuum of banks, $j \in [0, 1]$. All banks perform maturity transformation in the spirit of Diamond and Dybvig (1983). In other words, all banks have short-term liabilities and long-term assets at $t = 0$. Bank liabilities include

the overnight repo rate and lending by big banks to other SMBs was -0.62 while the correlation between the overnight repo rate and lending by policy banks to other SMBs was only -0.22. The much stronger correlation between the interbank repo rate and interbank lending by big banks suggests that big banks do indeed have the capacity to manipulate interbank prices.

deposits and WMPs. A dollar deposited at $t=0$ becomes $1+i_B$ if withdrawn at $t=1$ and $(1+i_B)^2$ 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 $t=2$. Bank assets include reserves and loans. A dollar lent at $t=0$ repays $(1+i_A)^2$ at $t=2$. Loans cannot be liquidated at $t=1$ so reserves are used to honor early withdrawals. For simplicity, we assume i_A and i_B are fixed with $i_A > i_B$. One can interpret i_B as the deposit rate ceiling and i_A as a loan rate floor, both of which bind with sufficient competition.[26]

Savings Households have one unit of savings which must be split between deposits and WMPs at $t=0$. Let $D_j(\xi_j)$ and $W_j(\xi_j)$ denote the demands for bank j's deposits and WMPs respectively. We assume $W_j(0)=0$ which means WMPs are only bought if they pay more interest than regular deposits. We also assume $W_j'(\cdot)>0$ and $W_j''(\cdot)\le 0$. In words, WMP demand increases with the amount of interest paid but the increase is not exponential. Moreover, WMP demand is a continuous function: deposits have an (unmodelled) convenience value which stops households from switching entirely to WMPs once $\xi_j>0$.

Market for Reserves Each bank can be in one of two states at $t=1$. The first is a low withdrawal state where fraction θ_ℓ of households withdraw their deposits and WMPs. The second is a high withdrawal state with fraction $\theta_h > \theta_\ell$. The low withdrawal state occurs with probability $\pi \in (0,1)$, making the expected withdrawal fraction $\overline{\theta} \equiv \pi\theta_\ell + (1-\pi)\theta_h$. To cover withdrawals, the banking system needs reserves. Denote bank j's reserve holdings by R_j. Bank j thus attracts household savings $D_j(\cdot)+W_j(\cdot)$ at $t=0$, holds R_j as reserves, and lends the rest. If R_j proves insufficient to cover bank j's withdrawals at $t=1$, then j borrows from an interbank market at interest rate i_L. Interbank lenders are banks with surplus reserves. We also allow for a supply of external funds, $\Psi(i_L) \equiv \psi(i_L - i_B)$, where

[26] On the loan rate, we abstract from borrower characteristics, namely 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). While political pressure has waned over the past decade, the private sector remains thorny. In this regard, one can imagine similar "risk-adjusted" returns from different borrowers.

$\psi \geq 0$. External funds can be interpreted as liquidity injections by the central bank.

Regulation The government imposes a 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 liabilities must be at least $\alpha \in (0,1)$. Given the structure of our model, reserves are only needed in $t=1$ so enforcement of the liquidity rule is confined to $t=0$. The relevant liabilities are deposits and WMPs. Whereas deposits must be booked on balance sheet, banks can choose where to manage WMPs and the loans 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:

$$\lambda_j\left(\xi_j,\tau_j,R_j\right) \equiv \frac{R_j}{D_j\left(\xi_j\right)+\left(1-\tau_j\right)W_j\left(\xi_j\right)} \geq \alpha \tag{1}$$

Use of off balance sheet vehicles is regulatory arbitrage as defined in Adrian et al (2013).[27] Note that $1-\lambda_j(\cdot)$ is our model's counterpart to the loan-to-deposit ratio in Section 3.

4.2 Results with Homogeneous Banks

Suppose total savings attracted by bank j take the form $D_j(\cdot)+W_j(\cdot) \equiv \rho_0+\rho_1 W_j(\cdot)$, where $\rho_0 > 0$ and $\rho_1 \in [0,1]$ are constants. This specification nests two special cases. If $\rho_1 = 0$, then bank j attracts a fixed amount of savings ρ_0: any WMPs it issues will cut one-for-one into its own deposit base. If $\rho_1 = 1$, then bank j attracts a fixed amount of deposits ρ_0: any WMPs it issues will cut into the savings available for other banks. Household savings are still normalized to one so (given all other parameters) a symmetric equilibrium requires ρ_0 that generates an optimal choice ξ_j^* satisfying $D_j\left(\xi_j^*\right)+W_j\left(\xi_j^*\right)=1$.

A symmetric equilibrium also requires interbank market clearing. Mathematically:

$$R_j^* + \Psi\left(i_L^*\right) = \overline{\theta}\left(1+i_B\right) \tag{2}$$

[27] Their definition is "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."

The left-hand side of equation (2) is total liquidity available at $t = 1$, namely bank reserves, R_j, and external liquidity, $\Psi(i_L)$. The right-hand side is total liquidity required, namely the sum of all household withdrawals. In a symmetric equilibrium, each bank attracts a unit of household savings at $t = 0$ and is thus liable for $1 + i_B$ at $t = 1$. With an average of $\overline{\theta}$ households withdrawing at $t = 1$, the banking system needs liquidity $\overline{\theta}(1 + i_B)$. The value of i_L that solves equation (2) is the equilibrium interbank rate. This rate clearly enters (2) through $\Psi(\cdot)$ but it can also enter indirectly through the optimal choice of R_j derived next.

The representative bank chooses the attractiveness of its WMPs ξ_j, the intensity of its off balance sheet activities τ_j, and its reserve holdings R_j to maximize expected profit at $t = 0$. The bank's choices must also satisfy the liquidity rule set out in (1), namely $\lambda_j(\cdot) \geq \alpha$, so we can write the optimization problem as:

$$\max_{\xi_j,\tau_j,R_j} \left\{ \begin{array}{l} (1+i_A)^2 \left[D_j\left(\xi_j\right) + W_j\left(\xi_j\right) - R_j\right] \\ + (1+i_L)\left[R_j - \overline{\theta}(1+i_B)\left[D_j\left(\xi_j\right) + W_j\left(\xi_j\right)\right]\right] \\ - \left(1-\overline{\theta}\right)\left[(1+i_B)^2\left[D_j\left(\xi_j\right) + W_j\left(\xi_j\right)\right] + \xi_j W_j\left(\xi_j\right)\right] \end{array} \right\} \tag{3}$$

$$\textit{subject to}$$

$$\lambda_j\left(\xi_j, \tau_j, R_j\right) \geq \alpha$$

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

The Lagrange multiplier on $\lambda_j(\cdot) \geq \alpha$ 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:

$$\mu_j = (1+i_A)^2 - (1+i_L) \tag{4}$$

$$\eta_j^1 = \eta_j^0 + \alpha\mu_j W_j\left(\xi_j\right) \tag{5}$$

$$\xi_j + \frac{W_j\left(\xi_j\right)}{W_j'\left(\xi_j\right)} = \rho_1\left[\frac{\left[1-\overline{\theta}(1+i_B)\right](1+i_A)^2}{1-\overline{\theta}} - (1+i_B)^2\right] - \frac{\rho_1\left[\alpha - \overline{\theta}(1+i_B)\right] - \alpha\tau_j}{1-\overline{\theta}}\mu_j \tag{6}$$

If $1 + i_L = (1 + i_A)^2$, then $\mu_j = 0$ from (4) and $\eta_j^1 = \eta_j^0$ from (5). In words, the liquidity rule is not binding and the bank is indifferent between any off balance sheet intensity $\tau_j \in [0, 1]$. Substituting $\mu_j = 0$ into equation (6) then reveals that the optimal choice of ξ_j hinges on ρ_1.[28] With $\rho_1 = 0$, bank j's WMPs cut one-for-one into its own deposit base. Deposits are a cheaper liability than WMPs so $\rho_1 = 0$ implies $\xi_j^* = 0$ and $W_j\left(\xi_j^*\right) = 0$. With $\rho_1 > 0$, the cut into bank j's deposits is only partial: the rest comes from savings available to other banks. This prompts $\xi_j^* > 0$ and $W_j\left(\xi_j^*\right) > 0$ but with $\frac{\partial \xi_j^*}{\partial \alpha} = 0$. Since the bank is indifferent between any $\tau_j \in [0, 1]$, WMP issuance under $\rho_1 > 0$ stems from competition for a larger share of household savings, not from a desire to evade liquidity requirements.

If $1 + i_L < (1 + i_A)^2$, then $\mu_j > 0$ and the liquidity rule binds. WMP issuance now arises to evade liquidity requirements. To see this, consider $\rho_1 = 0$ which previously resulted in no WMPs. Suppose the bank chooses $\tau_j^* = 1$. Equation (6) then returns $\xi_j^* > 0$ which, when substituted into (5), implies $\eta_j^1 > 0$ and confirms the choice of $\tau_j^* = 1$.[29] Notice that the incentive to issue WMPs no longer comes from competition: with $\rho_1 = 0$, 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.

It now remains to check how the equilibrium value of $1 + i_L$ compares to $(1 + i_A)^2$. The results are summarized in the next two propositions:

Proposition 1 *Suppose $\rho_1 = 0$ so regulatory arbitrage is the only motive for issuing WMPs. If $W_j(\cdot)$ is sufficiently concave, then there exists a scalar $\overline{\alpha} \in [0, 1]$ such that:*

1. $1 + i_L^* = (1 + i_A)^2$ *and* $\xi_j^* = 0$ *with* $\lambda_j^* = \overline{\alpha}$ *for any* $\alpha \leq \overline{\alpha}$
2. $1 + i_L^* < (1 + i_A)^2$ *and* $\xi_j^* > 0$ *with* $\tau_j^* = 1$, $\frac{di_L^*}{d\alpha} < 0$, $\frac{d\xi_j^*}{d\alpha} > 0$, *and* $\lambda_j^* = \alpha$ *for any* $\alpha > \overline{\alpha}$

[28] Assume the coefficient on ρ_1 in the first term on the right-hand side of equation (6) is positive so that a non-trivial solution exists.

[29] In principle, there is also a solution with $\xi_j = 0$: using $\tau_j = 0$ in (6) gives $\xi_j = 0$ which, when substituted into (5), is consistent with any $\tau_j \in [0, 1]$. However, we cannot go from any $\tau_j \in [0, 1]$ to $\tau_j = 0$ so the solution with $\xi_j = 0$ is eliminated by refinement.

In words, Proposition 1 says that sufficiently stricter regulation (i.e., increasing α from something below $\overline{\alpha}$ to something above $\overline{\alpha}$) triggers the issuance of off balance sheet WMPs and leads to a lower interbank rate. The model with homogeneous banks thus accounts for the rise of shadow banking but not tighter interbank conditions. As shown in Proposition 2 below, this shortcoming is not an artifact of $\rho_1 = 0$:

Proposition 2 *If $\rho_1 \in (0,1]$, then ψ sufficiently low ensures $D_j\left(\xi_j^*\right) > 0$. There is also an $\overline{\overline{\alpha}} \in [0,1]$ such that $1 + i_L^* = (1+i_A)^2$ if $\alpha \leq \overline{\overline{\alpha}}$ and $1 + i_L^* < (1+i_A)^2$ with $\frac{di_L^*}{d\alpha} < 0$ if $\alpha > \overline{\overline{\alpha}}$.*

We have now established that an increase in α leads to a lower interbank rate when banks are homogeneous. Given equation (2), total credit must also be lower. To see why, note that the total amount lent at $t = 0$ is $1 - R_j$ because household savings are normalized to one and banks hold R_j in reserves. With lower i_L, less external liquidity $\Psi(\cdot)$ is available to satisfy the same withdrawals. Banks must therefore hold more of their own reserves, prompting a fall in $1 - R_j$. If we were to eliminate $\Psi(\cdot)$, then the total amount lent at $t = 0$ would be a constant $1 - \overline{\theta}(1 + i_B)$ for any α. Either way, the simple model outlined here cannot generate more credit in the midst of tightening liquidity rules. By virtue of focusing on a representative bank, it is also silent on different responses by big versus small banks: the most we can glean from Proposition 1 is that small banks become constrained by the new liquidity rules, their loan-to-deposit ratio falls, and they move into off balance sheet WMPs.

4.3 Adding Heterogeneity in Market Power

Based on the discussion above, we now introduce a big bank. By definition of being big, the big bank does *not* take the interbank rate (or any other equilibrium object) as given.[30]

[30] Suppose everyone is a price-taker. Then, unless α is large, we need $1 + E(i_L) = (1+i_A)^2$ for reserve holdings to be sufficient at $t = 1$. With $1 + E(i_L) = (1+i_A)^2$, everyone is indifferent between holding reserves and lending. This is consistent with the argument in Farhi et al (2009). Given indifference, it is then possible that the small bank loan-to-deposit ratio is exactly $1 - \alpha$ while the big bank ratio is below $1 - \alpha$ but this is only one of many possibilities. It is also possible that an increase in α leads to convergence in these ratios but, again, there are many other possibilities. In short, an objective equilibrium selection criterion is missing: the model with everyone being a price-taker does not provide clear microfoundations for

We keep the continuum of small banks, $j \in [0,1]$, and index the big bank by k. Since the big bank is effectively a price-setter in the interbank market, the interbank rate will depend on the big bank's withdrawal fraction. We can thus interpret the big bank's individual state as an aggregate state. In particular, suppose k is hit by withdrawal fraction θ_s where $s \in \{\ell, h\}$. Then interbank clearing in state s requires:

$$R_j + R_k + \Psi\left(i_L^s\right) = \overline{\theta}\left(1 + i_B\right)\left(D_j + W_j\right) + \theta_s\left(1 + i_B\right)\left(D_k + W_k\right)$$

There are two important points here. First, all choices are made ex ante so the market can only clear in one state if $\Psi(\cdot)$ is not assumed to be very potent. If the market clears at i_L^h, then there is an excess supply of reserves at i_L^ℓ. If the market clears at i_L^ℓ, then there is excess demand for reserves at i_L^h. We do not assume $\Psi(\cdot)$ very potent. Instead, we assume the market clears at i_L^h and set $i_L^\ell = i_B$ to reduce notation.[31] The second important point is that the amount of liquidity needed at $t = 1$ is endogenous. In particular, it depends on the split between $D_j + W_j$ and $D_k + W_k$ which will itself depend on ξ_j and ξ_k. This differs from the homogenous model where the right-hand side of equation (2) was constant.

To fix ideas, we use the following functional forms for WMP demand:

$$W_j = \beta\xi_j\left(\xi_j + \xi_k\right)^{\gamma-1} \;\; and \;\; W_k = \beta\xi_k\left(\xi_j + \xi_k\right)^{\gamma-1}$$

where $\beta > 0$ and $\gamma \in [0,1]$. If $\gamma = 1$, then the demand for big bank WMPs only responds to ξ_k. If $\gamma = 0$, then it only responds to the relative value ξ_k/ξ_j. The demand for small bank WMPs depends on γ in an analogous way. Any household savings not allocated to WMPs are then divided between banks as traditional deposits. Mathematically, $\delta \in (0,1)$ with:

$$D_j = \delta\left[1 - \beta\left(\xi_j + \xi_k\right)^{\gamma}\right] \;\; and \;\; D_k = (1-\delta)\left[1 - \beta\left(\xi_j + \xi_k\right)^{\gamma}\right]$$

converging loan-to-deposit ratios. Such a model, by virtue of being stuck at $1 + E\left(i_L\right) = \left(1 + i_A\right)^2$, would also not generate a change in the interbank rate after an increase in α.

[31] One can interpret $i_L^\ell > 0$ as interest on reserves at the end of $t = 1$.

Small banks still solve (3) but using the functional forms here along with the expected interbank rate, $1 + \pi i_B + (1 - \pi) i_L^h$, in place of $1 + i_L$. The big bank solves:

$$\max_{\xi_k, \tau_k, R_k} \left\{ \begin{array}{l} (1 + i_A)^2 (D_k + W_k - R_k) \\ + \left[1 + \pi i_B + (1 - \pi) i_L^h\right] \left[R_k - \overline{\theta} (1 + i_B) (D_k + W_k)\right] \\ - \left(1 - \overline{\theta}\right) \left[(1 + i_B)^2 (D_k + W_k) + \xi_k W_k\right] \\ -\pi (1 - \pi) (\theta_h - \theta_\ell) (1 + i_B) \left(i_L^h - i_B\right) (D_k + W_k) \end{array} \right\}$$

$$\textit{subject to}$$

$$\lambda_k \geq \alpha \textit{ and } \tau_k \in [0, 1]$$

$$\xi_j = \xi_j (\xi_k, \tau_k, R_k) \textit{ and } i_L^h = i_L^h (\xi_k, \tau_k, R_k)$$

There are two differences relative to the small bank problem. First is the extra term $\pi (1 - \pi) (\theta_h - \theta_\ell) (1 + i_B) \left(i_L^h - i_B\right) (D_k + W_k)$ subtracted from the big bank's objective function. This arises because the big bank's state is the aggregate state: if the big bank gets hit by a high withdrawal shock, then it has to pay the higher interest rate on a higher amount of interbank borrowing. Second, and most importantly, is that the big bank internalizes how ξ_j and i_L^h depend on its own choices. The expression for ξ_j comes from the first order conditions of the small banks while i_L^h comes from interbank clearing. Therefore, in our model, big banks can be unconstrained by the loan-to-deposit cap because they internalize that more reserves reduce the cost of interbank borrowing.

4.4 Results with Heterogeneous Banks

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.[32] Given the

[32]Something to note here is that the model will not go from $\xi_j = 0$ to $\xi_j > 0$ when α goes from a low value to a high value. This is true even if small banks are not constrained by the loan limit (i.e., even if $\mu_j = 0$). Reminiscent of the simple model with $\rho_1 > 0$: regardless of α, there is a motive for WMP issuance that stems from competition for a larger share of household savings. Adding an extra cost of WMP issuance would neutralize some of the competitive motive and "normalize" the model so that the competition benefit of WMPs is offset by the extra cost when liquidity regulations are mild. In short, we would just need to

empirical facts in Section 3, we seek parameters for which only small banks are constrained by the loan-to-deposit cap. The results are summarized by the gray area in Figure 17 (see Appendix B for derivations). In particular, we get that only small banks are constrained if α is not too high and ψ is between some positive bounds. Recall that $1 - \alpha$ is the loan-to-deposit cap while ψ is the responsiveness of external liquidity to the interbank rate. If α is very high, then regulation is overly strict and everyone is constrained. If ψ is very low, then the interbank rate is so high that small banks choose to hold additional liquidity and are thus not constrained. If ψ is very high, then the big bank has insufficient influence on the interbank market, prompting it to behave like a small bank and hit the cap.

We now want to show that our model is capable of generating the other key facts in Section 3. This is also done in Figure 17. In particular, for any α and ψ inside the shaded red area, a stricter loan-to-deposit cap (i.e., an increase in α) leads to: (i) a decrease in the small bank loan-to-deposit ratio; (ii) an increase in the big bank loan-to-deposit ratio; (iii) an increase in the interbank interest rate; (iv) an increase in total credit; and (v) an increase in the fraction of credit extended off balance sheet.

The intuition begins as follows. Small banks move (more heavily) 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 household savings from the big bank. Recall that the big bank internalizes the benefit of reserve holdings to the interbank market. Therefore, compared to small banks, it makes fewer loans at $t = 0$ per unit of savings attracted. The reallocation of savings 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 household savings. 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

shift the first order condition for ξ_j down so that it starts close to zero for α low.

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 loans at $t = 0$. Notice that the big bank's strategy also contributes directly to the rise in its loan-to-deposit ratio.

In support of our intuition, we find that the big bank is much less responsive to changes in α when its market share, $D_k + W_k$, is held constant. We also find that our qualitative results are largely unchanged if k's choices are derived as a Nash equilibrium between two big players. Lastly, we find that WMP activity intensifies when the big bank is less able to manipulate the interbank market. To better understand the last point, suppose the central bank becomes more responsive to interbank market conditions. We can capture this with an increase in ψ. If the big bank now tries to increase i_L^h to defend its market share, a lot of external liquidity will enter and prevent the higher i_L^h from being an equilibrium. Increasing ψ thus decreases the big bank's control over the interbank market. For all parameter combinations inside the shaded red area in Figure 17, an increase in ψ leads to increases in both ξ_j and $\frac{\xi_k}{\xi_j}$. In words, big and small banks all offer higher WMP returns, with the big bank competing more intensely now that it is less able to manipulate the interbank market.

4.5 Placements Extension

We now discuss a simple extension that allows for placements between counterparts. Recall from Section 2 that placements are becoming a popular way to deal with CBRC's recent crackdown on bank-trust cooperation. Funds still flow from WMPs to trusts but an intermediary emerges to split the transaction into individually compliant parts (see Figure 5). We saw evidence that China Merchants Bank is intermediating in this way for other banks.

Industrial Bank – another SMB – is also a well-known issuer of placements and is generally thought to have innovated the practice well before CBRC's crackdown. Our objective for this section is thus two-fold. First, we want to explore why placements did not gain traction when Industrial first introduced them. Second, we want to understand how the growing popularity of placements will affect WMP activity going forward.

We have made two simplifying assumptions up to now. One was that loans and reserves are the only bank assets at $t = 0$. The other was that the loan-to-deposit cap $(1 - \alpha)$ and the reserve requirement (α) sum to one. Now suppose the reserve requirement is $\alpha_0 \in (0, \alpha)$ so there is a wedge. A slight wedge is realistic for China: reserve requirements have been at most 21.5% while the loan-to-deposit cap is 75%. Also suppose an intermediary – call it the placement bank – offers a new type of asset: interest-bearing placements. Loans are still the only fundamental source of return in our economy so placements cannot be more lucrative than loans. However, placements may be more lucrative than a reserve ratio above α_0.

In expectation, a dollar of reserves at $t = 0$ is worth one dollar if used at $t = 1$ and $1 + \pi i_B + (1 - \pi)\, i_L^h$ if held until $t = 2$, where returns between $t = 1$ and $t = 2$ come from the interbank market. Suppose the placement bank promises $1 + \pi i_B + (1 - \pi)\, i_L^h + \varepsilon$ between $t = 0$ and $t = 2$. Placements tend to be longer-term than normal (online) interbank loans so we model a placement as a two-period investment which cannot be liquidated at $t = 1$. To streamline the exposition, we also assume that $\varepsilon > 0$ is a constant. Of course, we only consider values of ε where the placement bank is solvent in equilibrium.[33]

Restrictions on bank-trust cooperation are introduced with a marginal cost. In particular, a bank which books fraction τ_j of its WMPs off balance sheet now faces a cost $c\tau_j$, where

[33] The way we model the interest rate on placements has two implications. First, the placement rate is positively correlated with the interbank rate. Second, the placement rate exceeds the interbank rate. The higher interest rate on placements can be interpreted as a liquidity premium while the positive correlation between rates is reasonable since the placement bank competes with the interbank market for funds. Indeed, between 2008 and 2014, the correlation between the placement rate paid by China Merchants Bank and the overnight repo rate was 0.81. CMB's placement rate also exceeded the repo rate by an average of 50 basis points, with the largest spreads occurring after CBRC began regulating bank-trust cooperation.

$c \geq 0$ is a constant. Setting $c = 0$ returns the model in Sections 4.3 and 4.4 subject to one caveat: the loan-to-deposit cap and the reserve requirement no longer sum to one so we must determine whether or not placements are actually held. Setting $c = \infty$ eliminates trusts and provides insight into what would happen after an outright ban on bank-trust cooperation.

Let x_j denote the placement decision of bank j (the notation is symmetric for k). At $t = 0$, bank j has non-reserve assets $D_j + W_j - R_j$. Fraction x_j of these assets are held as placements and fraction $1 - x_j$ are held as loans. If $\tau_j > 0$, then some of the loans are booked in an off balance sheet trust. To satisfy both the loan-to-deposit cap and the reserve requirement, the bank needs:

$$R_j \geq \max\left\{\alpha_0, \tfrac{\alpha - x_j}{1 - x_j}\right\}\left[D_j + (1 - \tau_j)\, W_j\right] \tag{7}$$

The first term in the *max*-operator is the reserve requirement. The second term is the complement of the loan-to-deposit cap. Placements do not count against the loan-to-deposit cap so higher x_j will reduce the stringency of this cap.

The results are summarized in Figure 18 (see Appendix C for derivations). As before, we focus on parameters where only small banks are constrained by regulation. For any parameters within the gray area of Figure 18, small banks move all WMPs off balance sheet and hold no placements when $c = 0$.[34] For any parameters within the dashed black area, trusts disappear and small banks hold placements when $c = \infty$. The intersection of these two areas is bounded in red. In other words, for any parameters within the red area, there exists a trust-only equilibrium when $c = 0$ and a placement-only equilibrium when $c = \infty$.

Before proceeding, let us briefly explain why a trust-only equilibrium is even possible. That is, why would a small bank constrained by (7) not use both trusts and placements to

[34] The gray area here is a subset of the gray area in Figure 17. In particular, the downward sloping boundary is the same while the (roughly) flat boundary moves up with ε. If ψ is low, then i_L^h is high and the return on placements is attractive unless ε is low. So, the higher is ε, the higher ψ must be to get $x_j = 0$.

reduce its reserves? The answer lies in the endogenous nature of i_L^h. If small banks hold very few reserves, then the expected interbank rate will be very high and small banks will want to increase their reserve holdings. In this regard, there is a floor on how many reserves are held in equilibrium. Near this floor, trusts and placements become substitutes and, in the gray area of Figure 18, the placement interest that small banks seemingly forgo is more than compensated by the interest earned on off balance sheet loans. Once restrictions on bank-trust cooperation are introduced, the return to going off balance sheet falls and placements become the optimal choice. Our extended model can thus account for the paucity of placements before CBRC's crackdown and the rise of placements afterwards.

Our extended model also suggests that a ban on bank-trust cooperation can actually increase WMP issuance. As indicated in Figure 18, the placement-only equilibrium involves lower i_L^h and higher $\frac{\xi_k}{\xi_j}$ than the trust-only equilibrium. In words, banning bank-trust cooperation prompts the big bank to set a lower interbank rate and compete directly on WMP returns. Moreover, for any parameters within the dashed blue area, the placement-only equilibrium also involves a lower loan-to-deposit ratio for the big bank and an increase in total WMPs outstanding. WMP returns offered by small banks (ξ_j) are also higher for any parameters in the blue area bounded above by the dotted line. Key to all these results is that interbank manipulation is less attractive to the big bank. This is because higher interbank rates now have two competing effects. The first effect is as discussed in Section 4.4: all else constant, higher i_L^h forces small banks to scale back their issuance of WMPs. The second effect arises because of placements. In particular, higher i_L^h decreases the return differential between loans and placements without changing the fact that placements help small banks alleviate (7). To put it another way, the average second period return of small banks who use placements increases with i_L^h. These banks can then offer higher returns on their WMPs and cut into the big bank's deposit base. Placements thus make it harder for the big bank to defend its deposits using i_L^h, forcing it to compete directly on WMP returns.

5 Conclusion

This paper has explored 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. Evading these rules allows banks to be very aggressive at attracting and intermediating household savings. On one hand, competition for savings increases. On the other, small and medium-sized banks rely on interbank markets for emergency liquidity, giving large interbank lenders cartel power. The cartel then begins favoring traditional lending over interbank lending to undermine the shadow banks rather than compete directly with them for savings. The net result is an increase in both traditional and shadow lending as well as an increase in interbank interest rates, making the stricter liquidity rules entirely counterproductive.

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.

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.

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.

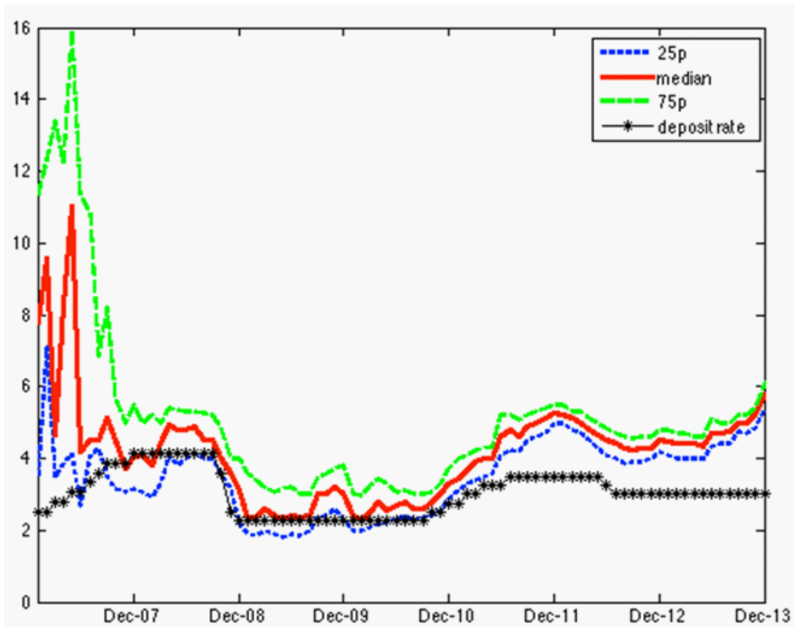
Krishnamurthy, A., S. Nagel, and D. Orlov. 2014. "Sizing Up Repo." *Journal of Finance*, 69(6), pp. 2381–2417.

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

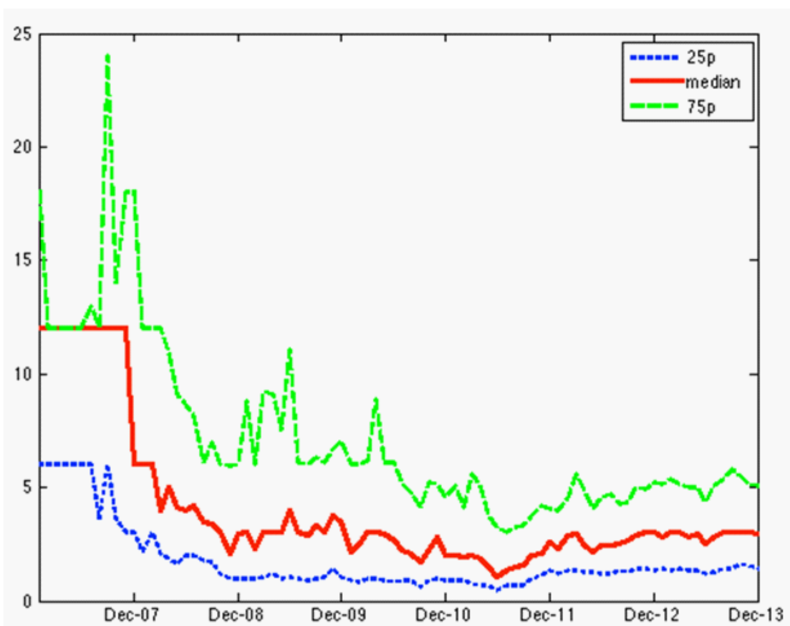
Song, Z., K. Storesletten, and F. Zilibotti. 2011. "Growing Like China." *American Economic Review*, 101(1), pp. 202-241.

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

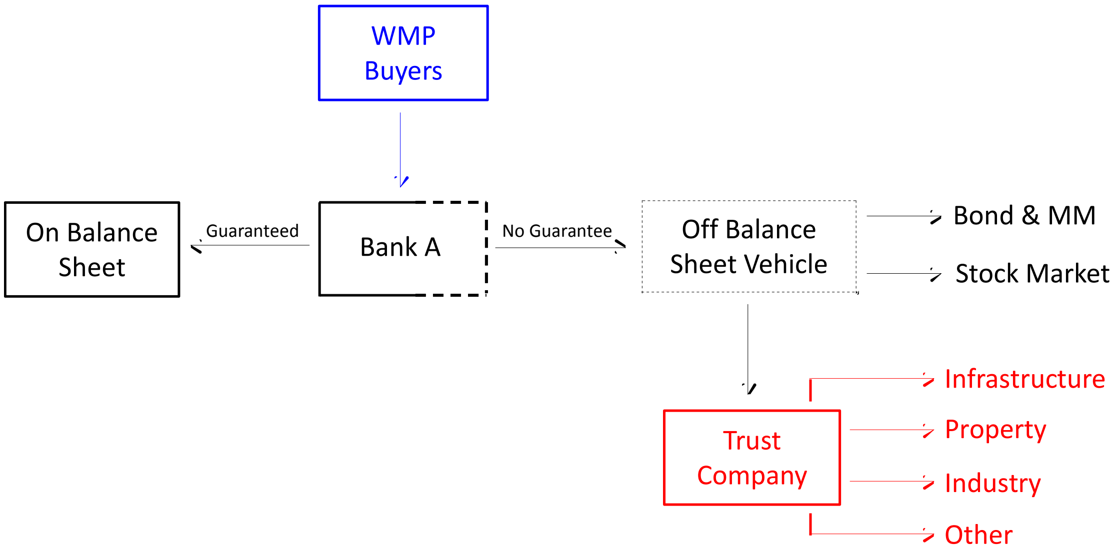
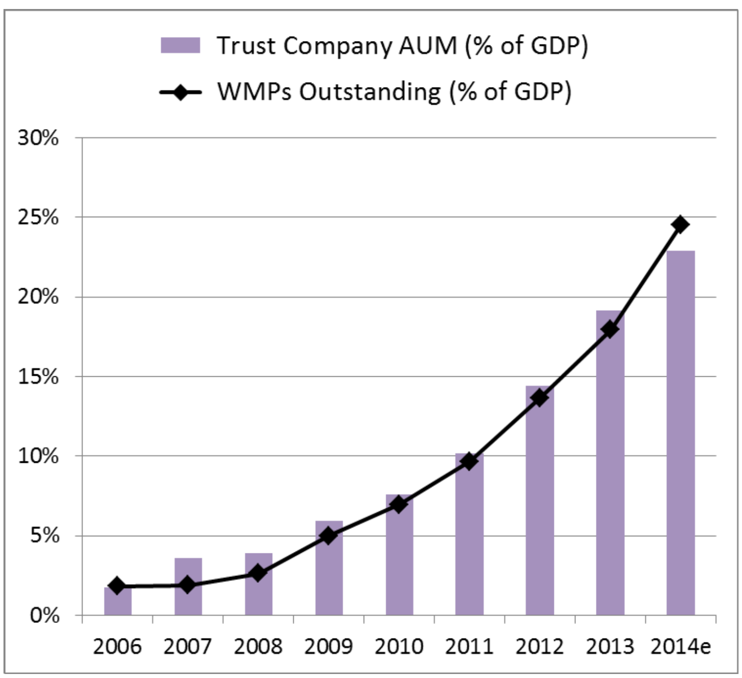
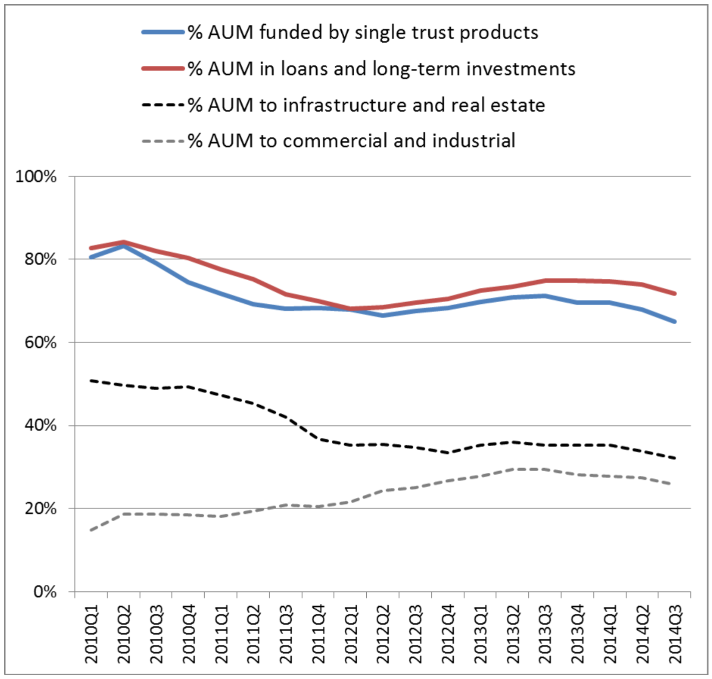


Figure 3



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

Figure 4



Source: China Trustee Association

Figure 5

Business with Counterparts

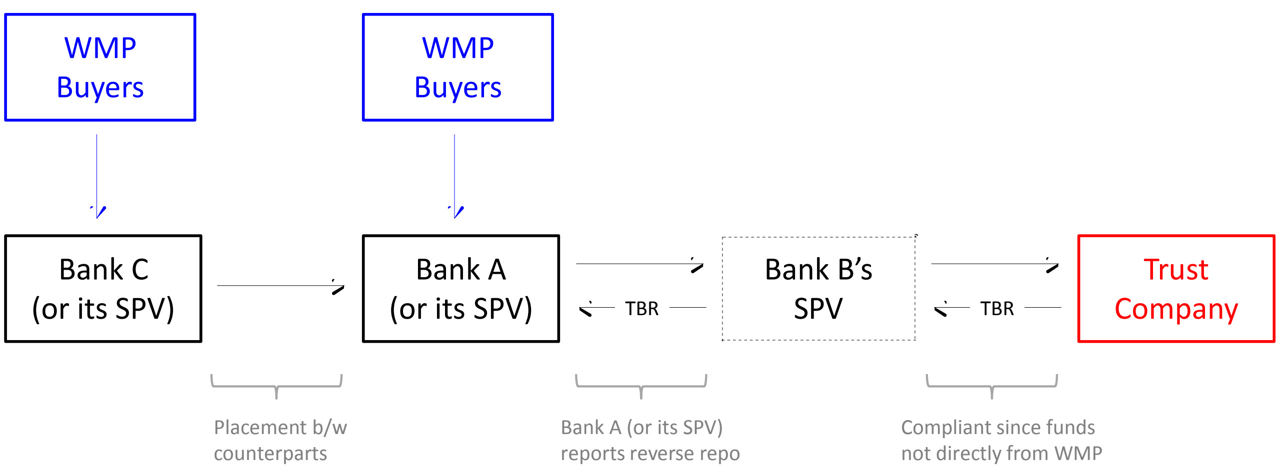
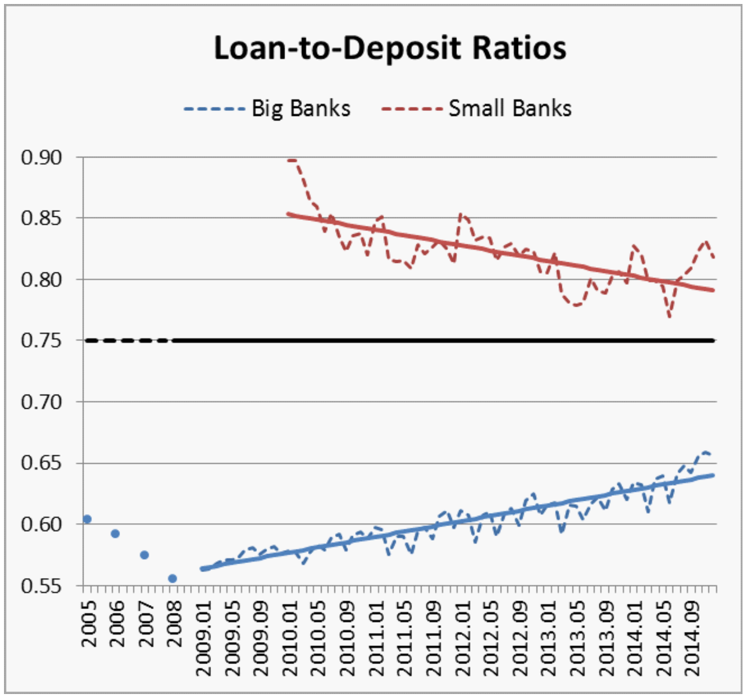
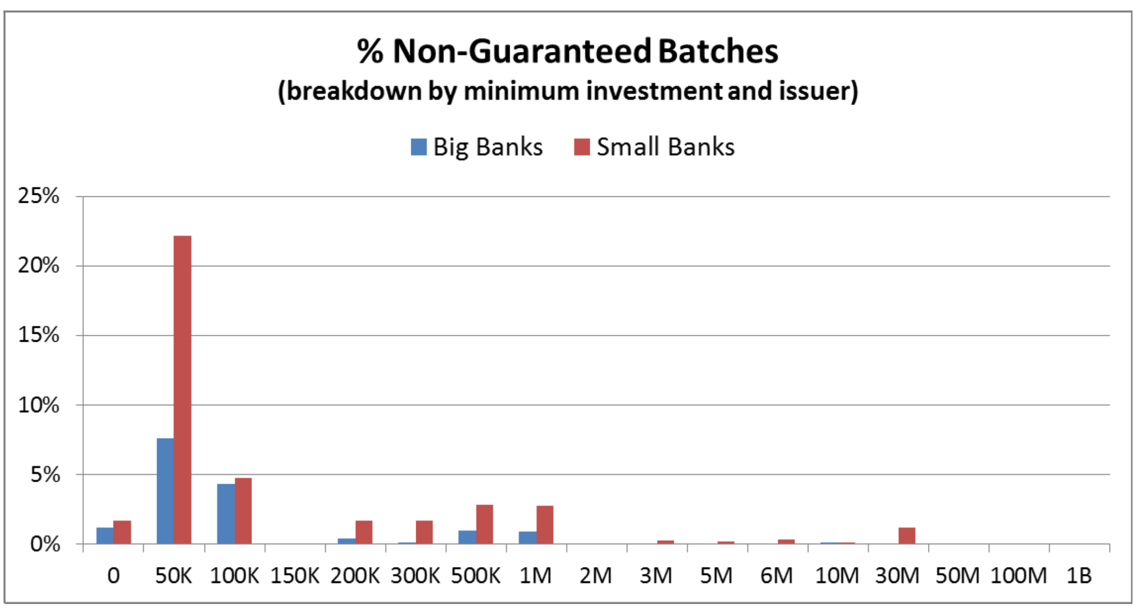


Figure 6



Source: PBOC. Pre-2009 data is from bank annual reports.

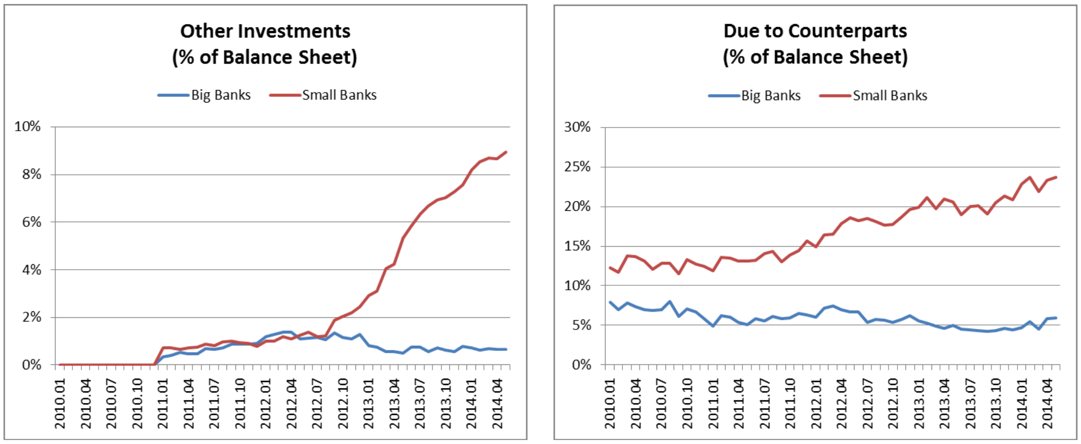
Figure 7



Source: China Banking Financial Network

Figure 8

Panel (a) Panel (b)



Source: PBOC (Financial Institutions Statistics)

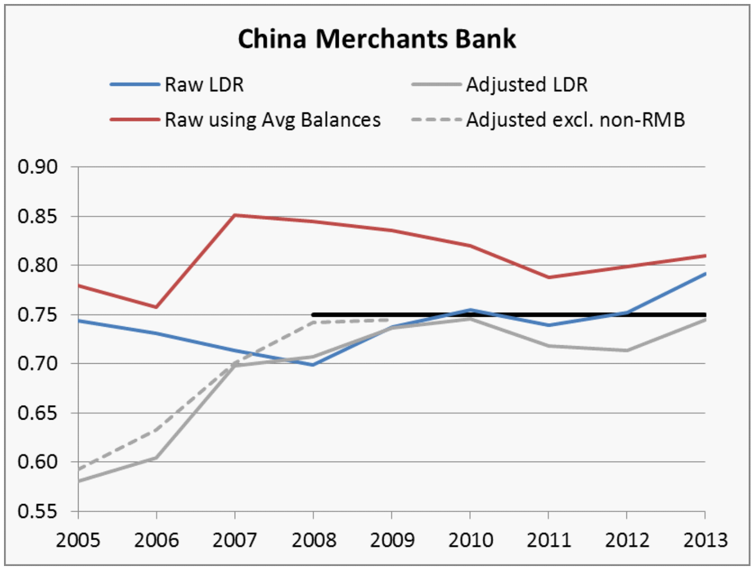
Figure 9

Granger Causality Wald Tests

	Chi2	Prob > Chi2
H0: # of WMP issued by the small banks does not Granger-cause # of WMP issued by the big four	21.104	**0.002**
H0: # of WMP issued by the big four does not Granger-cause # of WMP issued by the small banks	5.5264	**0.478**

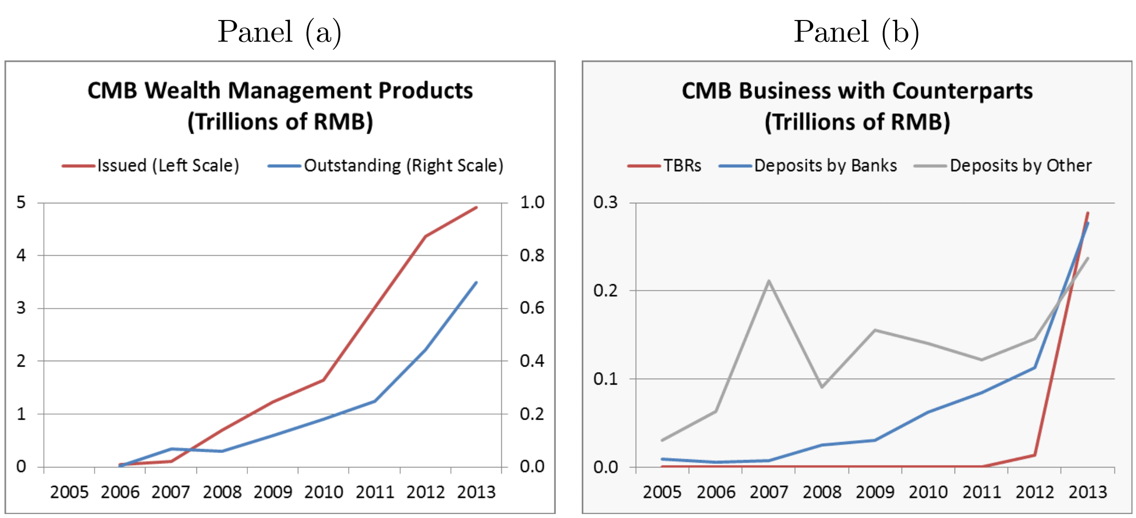
Note: We use detrended monthly data from Wind and estimate VARs with six lags

Figure 10



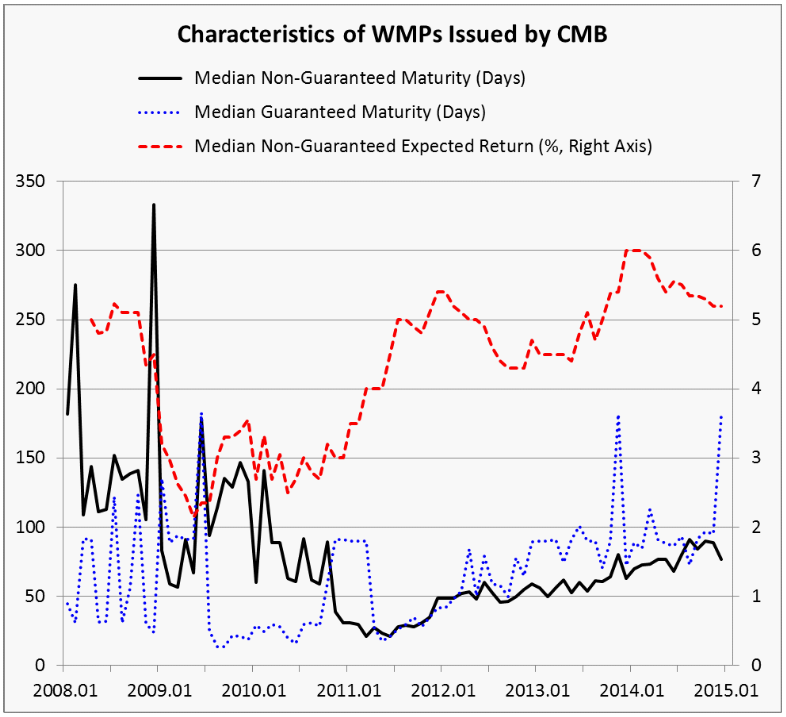
Source: China Merchants Bank Annual Reports

Figure 11



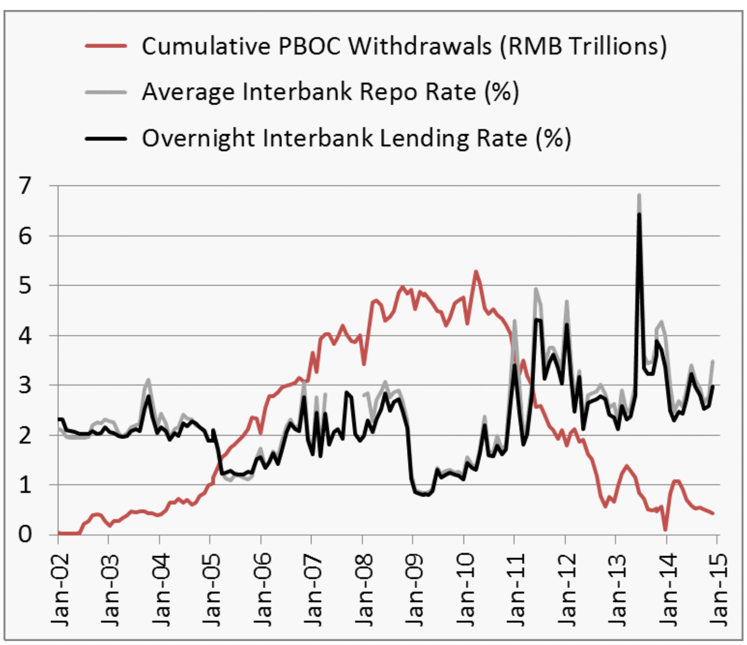
Source: China Merchants Bank Annual Reports

Figure 12



Source: Wind Financial Terminal

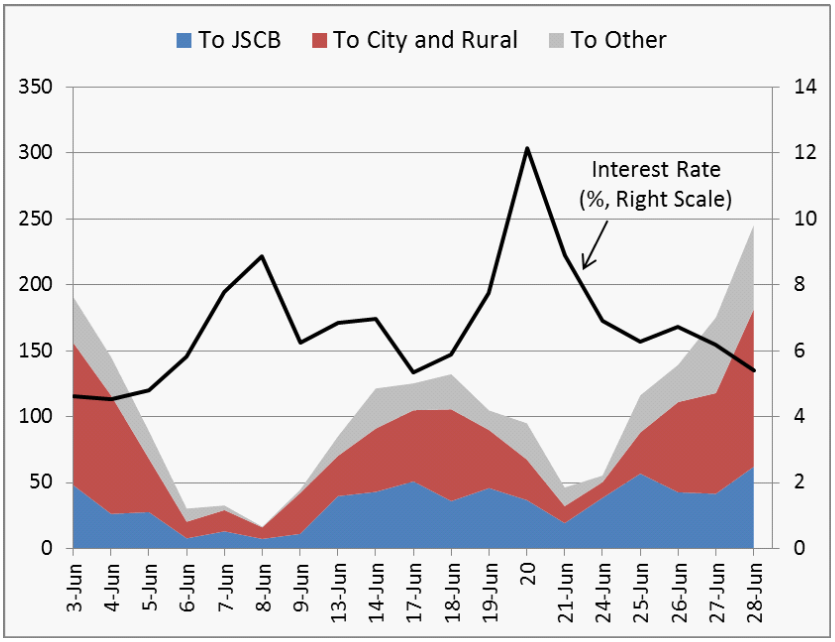
Figure 13



Source: PBOC and Wind Financial Terminal

Figure 14

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



Note: Excludes lending between big banks

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

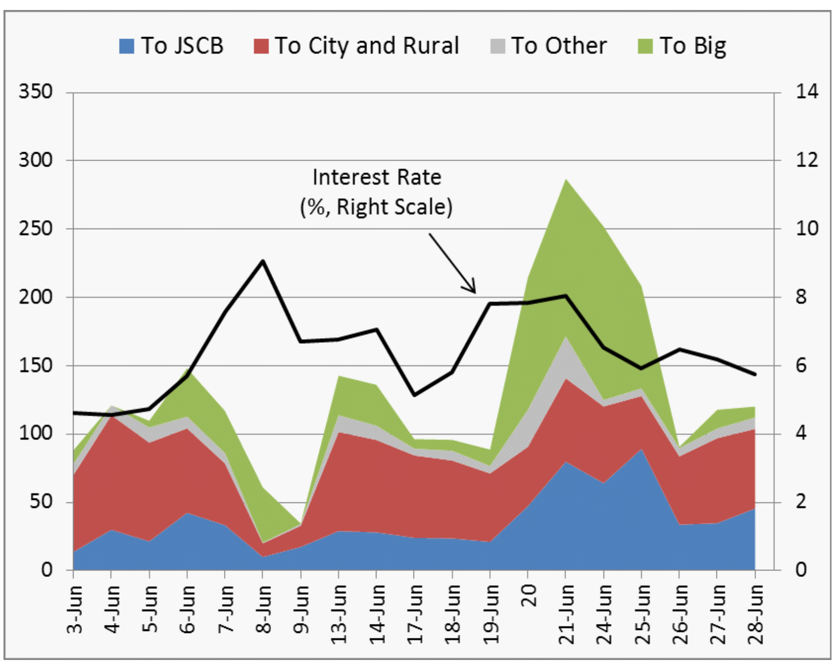


Figure 15

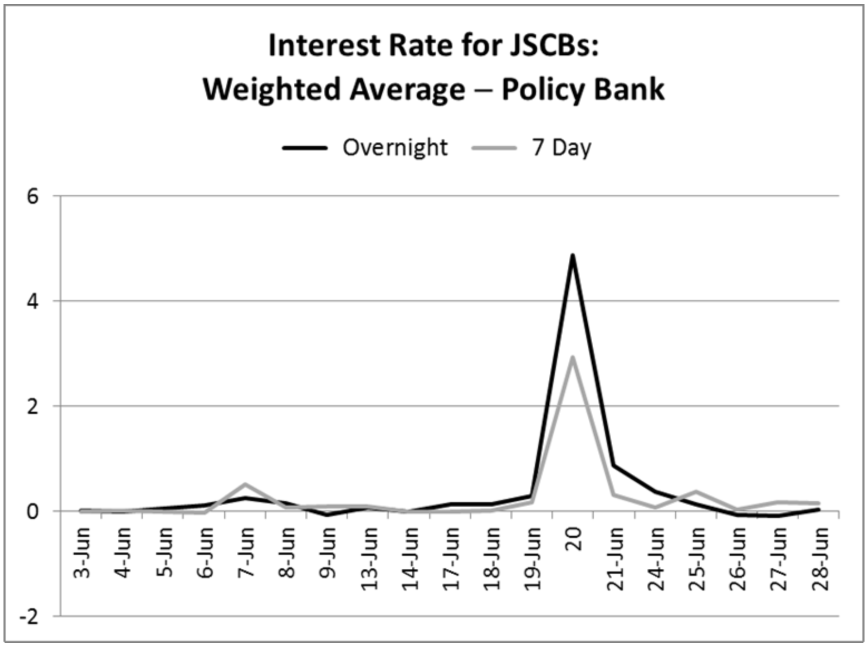


Figure 16

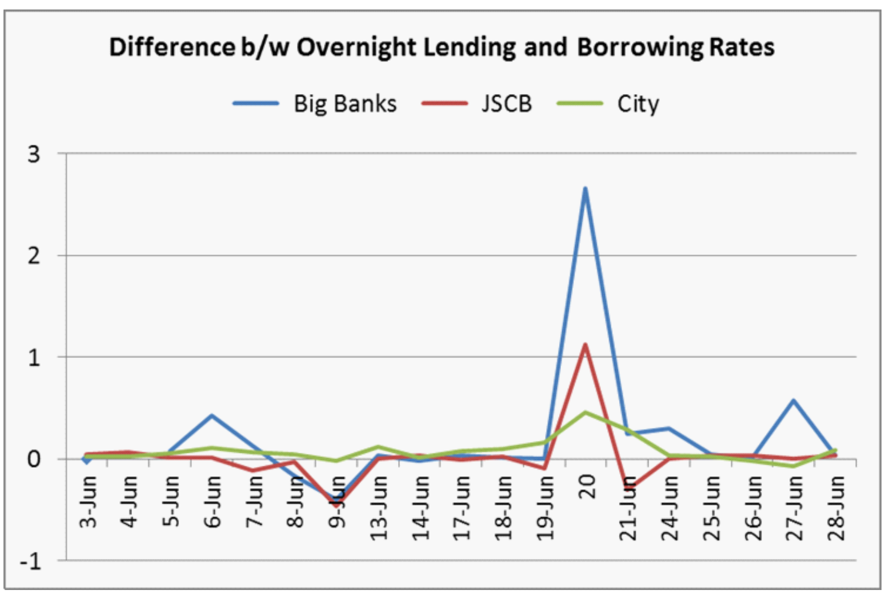
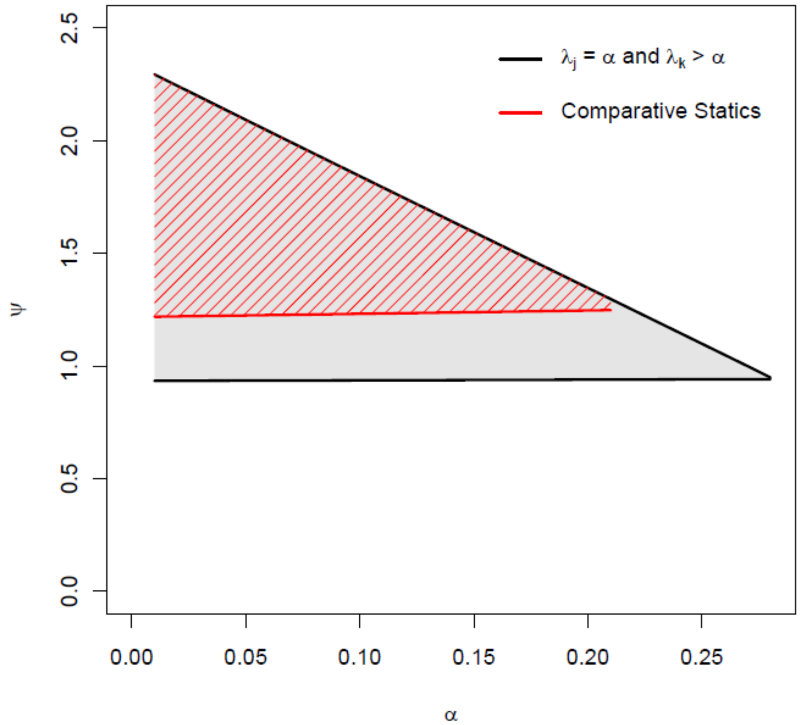


Figure 17

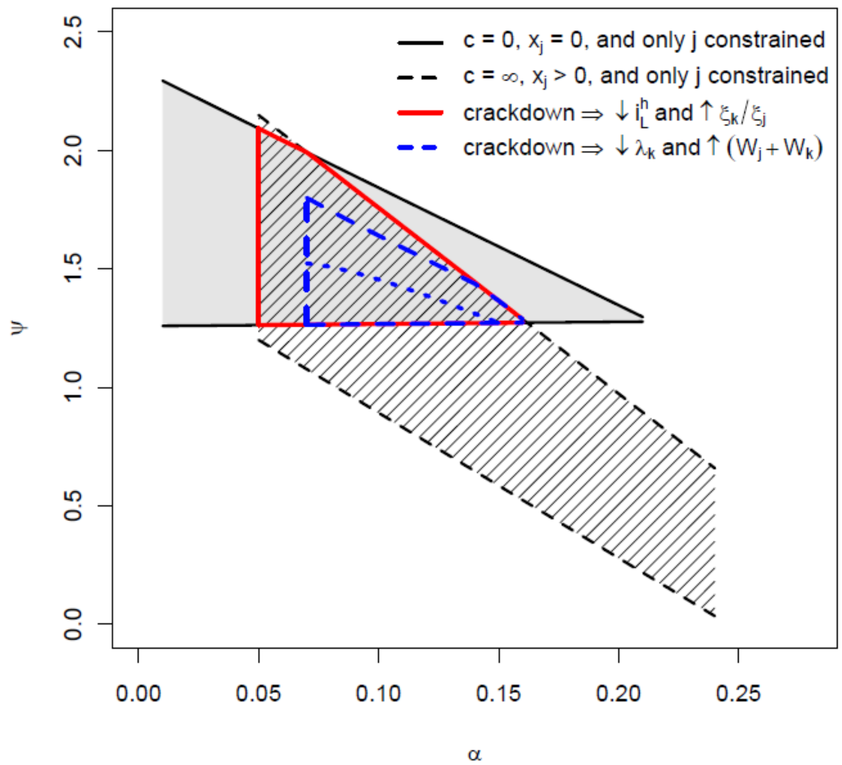
Parameter Space for Main Model



Generated with $i_A = \sqrt{1.1} - 1$, $i_B = 0.02$, $\theta_\ell = 0.05$, $\theta_h = 0.35$, $\pi = 0.5$, $\beta = 2$, $\gamma = 0.2$, $\delta = 0.3$

Figure 18

Parameter Space for Placements Extension



Other parameters as in Figure 17 with $\varepsilon = 0.01$ and $\alpha_0 = \alpha - 0.05$

Appendix A – Proofs for Section 4.2

Proof of Proposition 1

Begin with $1+i_L=(1+i_A)^2$. The main text already established $\xi_j=0$ and $W_j=0$. In equilibrium, $D_j+W_j=1$ so $D_j=1$ and the bank's liquidity rule is $R_j\geq\alpha$. Substituting $1+i_L=(1+i_A)^2$ into equation (2) pins down $R_j=\alpha_0\equiv\left(\psi+\overline{\theta}\right)(1+i_B)-\psi\,(1+i_A)^2$. For this to satisfy $R_j\geq\alpha$, we need $\alpha\leq\alpha_0$.

Consider now $1+i_L<(1+i_A)^2$. We already know $\tau_j=1$ so equations (4) and (6) give:

$$\xi_j+\frac{W_j\left(\xi_j\right)}{W_j'\left(\xi_j\right)}=\frac{\alpha\left[(1+i_A)^2-(1+i_L)\right]}{1-\overline{\theta}} \tag{8}$$

We also know that the liquidity rule binds so use $D_j+W_j=1$ to write $R_j=\alpha\,(1-W_j)$. We can now substitute into equation (2) to get another relationship between ξ_j and i_L:

$$i_L=i_B+\frac{\overline{\theta}(1+i_B)-\alpha\left[1-W_j\left(\xi_j\right)\right]}{\psi} \tag{9}$$

Totally differentiate equations (8) and (9) then combine to find:

$$\frac{di_L}{d\alpha}\overset{sign}{=}W_j\left(\xi_j\right)+\xi_jW_j'\left(\xi_j\right)-\left[1-W_j\left(\xi_j\right)\right]\left[2-\frac{W_j\left(\xi_j\right)}{W_j'\left(\xi_j\right)}\frac{W_j''\left(\xi_j\right)}{W_j'\left(\xi_j\right)}\right]$$

The sign of $\frac{di_L}{d\alpha}$ thus depends on the curvature of $W_j\left(\cdot\right)$. In particular, if $W_j\left(\cdot\right)$ is sufficiently concave, then $\frac{di_L}{d\alpha}<0$. Notice that $\frac{d\xi_j}{d\alpha}>0$ follows from equation (8). We now need to confirm $1+i_L<(1+i_A)^2$. Using equation (9), this requires $x\left(\alpha\right)\equiv\alpha\left[1-W_j\left(\xi_j\left(\alpha\right)\right)\right]>\alpha_0$. The same condition that yields $\frac{di_L}{d\alpha}<0$ also yields $x'\left(\alpha\right)>0$. Moreover, $x\left(\alpha_0\right)=\alpha_0$ in the non-trivial case of $\alpha_0>0$. Therefore, $x\left(\alpha\right)>\alpha_0$ for any $\alpha>\alpha_0$, confirming $1+i_L<(1+i_A)^2$.

It now follows that low α yields the $1+i_L=(1+i_A)^2$ equilibrium while high α yields the $1+i_L<(1+i_A)^2$ one. Defining $\overline{\alpha}\equiv\min\left\{\max\left\{\alpha_0,0\right\},1\right\}$ completes the proof. ■

Proof of Proposition 2

If $\rho_1 > 0$, then $W_j = \frac{1-\rho_0}{\rho_1}$ and $D_j = 1 - \frac{1-\rho_0}{\rho_1}$. Begin with $1 + i_L = (1 + i_A)^2$. Substitute i_L into the market clearing equation to get $R_j = \alpha_0$. The bank's liquidity rule is $R_j \geq \alpha\left[1 - \frac{(1-\rho_0)\tau_j}{\rho_1}\right]$ with $\tau_j \in [0,1]$ so we again need α below some threshold. Turn now to $1 + i_L < (1 + i_A)^2$. The liquidity rule binds with $\tau_j = 1$, implying $R_j = \alpha\left[1 - \frac{1-\rho_0}{\rho_1}\right]$. Substituting into market clearing yields:

$$i_L = i_B + \frac{\overline{\theta}(1+i_B)}{\psi} - \frac{\alpha}{\psi}\left[1 - \frac{1-\rho_0}{\rho_1}\right] \tag{10}$$

Confirming $1 + i_L < (1 + i_A)^2$ thus requires $\left[1 - \frac{1-\rho_0}{\rho_1}\right]\alpha > \alpha_0$. Moreover:

$$\frac{di_L}{d\alpha} = -\frac{1}{\psi}\left[1 - \frac{1-\rho_0}{\rho_1}\right]$$

This has to be negative otherwise $D_j \leq 0$ which we can rule out as follows. Notice that $D_j > 0$ requires $\rho_0 > 1 - \rho_1$. Also notice that equation (6) with $\tau_j = 1$ and i_L as per (10) must yield ξ_j consistent with $W_j\left(\xi_j\right) = \frac{1-\rho_0}{\rho_1}$. Such consistency amounts to a particular value of ρ_0 conditional on all other parameters so we just need to show that this particular value satisfies $\rho_0 > 1 - \rho_1$. Define:

$$\Delta_j \equiv \frac{\left[1-\overline{\theta}(1+i_B)\right](1+i_A)^2}{1-\overline{\theta}} - (1 + i_B)^2$$

Using (10), write equation (6) as:

$$\xi_j + \frac{W_j\left(\xi_j\right)}{W_j'\left(\xi_j\right)} = \rho_1\Delta_j + \frac{\alpha(1-\rho_1)+\overline{\theta}(1+i_B)\rho_1}{\psi\left(1-\overline{\theta}\right)}\left[\alpha\left(1 - \frac{1-\rho_0}{\rho_1}\right) - \alpha_0\right] \tag{11}$$

This expression implies $\frac{d\xi_j}{d\rho_0} > 0$ whereas $W_j\left(\xi_j\right) = \frac{1-\rho_0}{\rho_1}$ implies $\frac{d\xi_j}{d\rho_0} < 0$. So, all else constant, there is at most one value of ρ_0 that works. Substitute $\rho_0 = 1 - \rho_1$ into equation (11). The result is clearly impossible if the right-hand side is negative or, equivalently, if:

$$(1+i_A)^2 - (1+i_B)^2 + \alpha\left(\frac{1}{\rho_1} - 1\right)\left[(1+i_A)^2 - (1+i_B) - \frac{\overline{\theta}(1+i_B)}{\psi}\right] \leq \frac{\overline{\theta}^2(1+i_B)^2}{\psi}$$

A sufficient condition for the above inequality is:

$$\psi \leq \min\left\{\frac{\overline{\theta}(1+i_B)}{(1+i_A)^2-(1+i_B)}, \frac{\overline{\theta}^2(1+i_B)^2}{(1+i_A)^2-(1+i_B)^2}\right\} = \frac{\overline{\theta}^2(1+i_B)^2}{(1+i_A)^2-(1+i_B)^2}$$

where the solution to the min operation follows from $\Delta_j > 0$. ■

Appendix B – Derivations for Section 4.4

We first derive the equations that define an equilibrium where only small banks are constrained by the loan limit. We then find the parameter conditions to support this equilibrium. Defining some terms will streamline the exposition. First, define the following constants:

$$\Phi \equiv \alpha - \overline{\theta}\,(1+i_B)$$

$$\Theta \equiv \frac{\pi(\theta_h-\theta_\ell)(1+i_B)\left[(1+i_A)^2-(1+i_B)\right]}{1-\overline{\theta}}$$

$$\Delta_k \equiv \frac{[1-\theta_h(1+i_B)](1+i_A)^2-(1-\theta_h)(1+i_B)^2}{1-\overline{\theta}}$$

$$Y(\alpha) \equiv \frac{(1+i_A)^2-(1+i_B)}{1-\pi} + \frac{\delta\left[\alpha-\overline{\theta}(1+i_B)\right]}{\psi}$$

Next, define the ratio $z \equiv \xi_k/\xi_j$ and the following functions of z:

$$m(z) \equiv 2\gamma\,(\gamma+z)\,[1-\delta\,(1+z)] - (1+\gamma+2z)$$

$$n(z) \equiv \gamma\,(1+\gamma) + 4\gamma z + \left(1+\gamma^2\right) z^2$$

$$q(z) \equiv (\gamma+z)\left[1+\gamma+2z-(1-\gamma)\,z^2\right] + z\left[(1+\gamma)\,(1+2z)+2z^2\right]$$

Finally, define the following functions of i_L^h:

$$g\left(i_L^h\right) \equiv \frac{(1+i_A)^2-(1+i_B)^2}{1-\overline{\theta}} - \frac{\overline{\theta}(1-\pi)(1+i_B)}{1-\overline{\theta}}\left(i_L^h - i_B\right)$$

$$f\left(i_L^h\right) \equiv \frac{(1+i_A)^2-\left[1+\pi i_B+(1-\pi)i_L^h\right]}{1-\overline{\theta}}$$

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. Begin with the small bank. The first order conditions with respect to R_j and τ_j are:

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

$$\eta_j^1 = \eta_j^0 + \alpha\mu_j W_j$$

As before, $\mu_j > 0$ yields $R_j = \alpha D_j$.[35] The first order condition with respect to ξ_j is:

$$\xi_j + \frac{W_j}{\partial W_j/\partial \xi_j} = \left[\Delta_j - \frac{\alpha-\overline{\theta}(1+i_B)}{1-\overline{\theta}}\mu_j\right]\left[1+\frac{\partial D_j/\partial\xi_j}{\partial W_j/\partial\xi_j}\right] + \frac{\alpha\tau_j}{1-\overline{\theta}}\mu_j$$

If $\xi_j > 0$, then $\tau_j = 1$ and we can use the functional forms to write the choice of $\xi_j > 0$ as:

$$\left[1+\frac{\xi_k+\gamma\xi_j}{\xi_j+\xi_k}\right]\xi_j = \alpha\delta\gamma f\left(i_L^h\right) + \left[\Delta_j + \overline{\theta}\,(1+i_B)\, f\left(i_L^h\right)\right]\left[\frac{\xi_k+\gamma\xi_j}{\xi_j+\xi_k} - \delta\gamma\right] \tag{12}$$

Based on this expression, the reactions to the big bank's choices are:

$$\frac{d\xi_j}{d\xi_k} = \frac{\xi_j}{\xi_k + \frac{\left(\xi_j+\xi_k\right)\left[(1+\gamma)\xi_j+2\xi_k\right]}{(1-\gamma)\left[\Delta_j+\overline{\theta}(1+i_B)f\left(i_L^h\right)-\xi_j\right]}}$$

$$\frac{d\xi_j}{di_L^h} = -\frac{\frac{1-\pi}{1-\overline{\theta}}\left[\overline{\theta}(1+i_B)+\left[\alpha-\overline{\theta}(1+i_B)\right]\frac{\delta\gamma\left(\xi_j+\xi_k\right)}{\xi_k+\gamma\xi_j}\right]}{1+\frac{\xi_j+\xi_k}{\xi_k+\gamma\xi_j}+\left[\Delta_j+\overline{\theta}(1+i_B)f\left(i_L^h\right)-\xi_j\right]\frac{(1-\gamma)\xi_k}{\xi_j+\xi_k}\frac{1}{\xi_k+\gamma\xi_j}}$$

[35] More precisely, $\mu_j > 0$ yields $R_j = \alpha\left[D_j + (1-\tau_j)\, W_j\right]$ but we can show that $(1-\tau_j)\, W_j$ is always zero. If $W_j = 0$, then $(1-\tau_j)\, W_j = 0$ is trivially true. If $W_j > 0$, then $\eta_j^1 > 0$ and thus $\tau_j = 1$ so we again have $(1-\tau_j)\, W_j = 0$. Of course, the jump from $W_j > 0$ to $\tau_j = 1$ assumes $\alpha > 0$.

Turn next to the big bank (without the loan limit constraint). The first order condition with respect to ξ_k is:

$$\begin{aligned}\Delta_k\left[\frac{\partial(D_k+W_k)}{\partial\xi_k}+\frac{\partial(D_k+W_k)}{\partial\xi_j}\frac{\partial\xi_j}{\partial\xi_k}\right] &= W_k+\xi_k\left[\frac{\partial W_k}{\partial\xi_k}+\frac{\partial W_k}{\partial\xi_j}\frac{\partial\xi_j}{\partial\xi_k}\right]\\ &+f\left(i_L^h\right)\left[\overline{\theta}\left(1+i_B\right)\left[\frac{\partial(D_j+W_j)}{\partial\xi_k}+\frac{\partial(D_j+W_j)}{\partial\xi_j}\frac{\partial\xi_j}{\partial\xi_k}\right]-\alpha\left[\frac{\partial D_j}{\partial\xi_k}+\frac{\partial D_j}{\partial\xi_j}\frac{\partial\xi_j}{\partial\xi_k}\right]\right]\end{aligned}$$

The first order condition with respect to i_L^h is:

$$\begin{aligned}\Delta_k\frac{\partial(D_k+W_k)}{\partial\xi_j}\frac{\partial\xi_j}{\partial i_L^h} &= \xi_k\frac{\partial W_k}{\partial\xi_j}\frac{\partial\xi_j}{\partial i_L^h}+f\left(i_L^h\right)\left[\overline{\theta}\left(1+i_B\right)\frac{\partial(D_j+W_j)}{\partial\xi_j}-\alpha\frac{\partial D_j}{\partial\xi_j}\right]\frac{\partial\xi_j}{\partial i_L^h}\\ &+f'\left(i_L^h\right)\left[\overline{\theta}\left(1+i_B\right)\left(D_j+W_j\right)-\alpha D_j-\psi\left(i_L^h-i_B\right)\right]-\psi f\left(i_L^h\right)\end{aligned}$$

Using the functional forms, we can write these equations as:

$$\begin{aligned}&\left[\Delta_k+\overline{\theta}\left(1+i_B\right)f\left(i_L^h\right)\right]\left[\left[1-\gamma\left(1-\delta\right)\right]\xi_j+\gamma\delta\xi_k-\left[\left(1-\gamma\delta\right)\xi_k+\gamma\left(1-\delta\right)\xi_j\right]\frac{\partial\xi_j}{\partial\xi_k}\right]\\ =\ &\xi_k\left[2\xi_j+\left(1+\gamma\right)\xi_k-\left(1-\gamma\right)\xi_k\frac{\partial\xi_j}{\partial\xi_k}\right]+\alpha\gamma\delta\left[1+\frac{\partial\xi_j}{\partial\xi_k}\right]\left(\xi_j+\xi_k\right)f\left(i_L^h\right)\end{aligned}$$

$$\begin{aligned}&\left[\Delta_k+\overline{\theta}\left(1+i_B\right)f\left(i_L^h\right)\right]\left[\left(1-\gamma\delta\right)\xi_k+\gamma\left(1-\delta\right)\xi_j\right]\frac{\partial\xi_j}{\partial i_L^h}\\ =\ &\left(1-\gamma\right)\xi_k^2\frac{\partial\xi_j}{\partial i_L^h}-\alpha\gamma\delta\frac{\partial\xi_j}{\partial i_L^h}\left(\xi_j+\xi_k\right)f\left(i_L^h\right)-f'\left(i_L^h\right)\frac{\overline{\theta}(1+i_B)(D_j+W_j)-\alpha D_j}{\beta\left(\xi_j+\xi_k\right)^{\gamma-2}}+\frac{\psi\left[f\left(i_L^h\right)+f'\left(i_L^h\right)\left(i_L^h-i_B\right)\right]}{\beta\left(\xi_j+\xi_k\right)^{\gamma-2}}\end{aligned}$$

Using $z\equiv\xi_k/\xi_j$, we can then rewrite as:

$$\frac{(1-\gamma)z^2}{1+z}\xi_j+\frac{\Delta_k+\overline{\theta}(1+i_B)f\left(i_L^h\right)-2z\xi_j}{1+\frac{\partial\xi_j}{\partial\xi_k}}=\alpha\gamma\delta f\left(i_L^h\right)+\left[\Delta_k+\overline{\theta}\left(1+i_B\right)f\left(i_L^h\right)\right]\left[\frac{\gamma+z}{1+z}-\gamma\delta\right]$$

$$\begin{aligned}&\left[\alpha\gamma\delta f\left(i_L^h\right)+\left[\Delta_k+\overline{\theta}\left(1+i_B\right)f\left(i_L^h\right)\right]\left[\frac{\gamma+z}{1+z}-\gamma\delta\right]\right]\frac{\partial\xi_j}{\partial i_L^h}\frac{1}{\xi_j(1+z)}\\ =\ &\frac{(1-\gamma)z^2}{(1+z)^2}\frac{\partial\xi_j}{\partial i_L^h}+\frac{\overline{\theta}(1+i_B)(1-\pi)}{1-\overline{\theta}}\frac{1}{1+z}+\frac{\delta(1-\pi)\left[\alpha-\overline{\theta}(1+i_B)\right]}{1-\overline{\theta}}+\frac{\psi\left[2f\left(i_L^h\right)-\frac{1-\pi}{1-\overline{\theta}}Y(\alpha)\right]}{\beta\xi_j^{\gamma}(1+z)^{\gamma}}\end{aligned}$$

Putting everything together, the equilibrium is summarized by:

$$\left[1+\tfrac{\gamma+z}{1+z}\right]\xi_j = \alpha\gamma\delta f\left(i_L^h\right) + \left[\Delta_j + \overline{\theta}\left(1+i_B\right) f\left(i_L^h\right)\right]\left[\tfrac{\gamma+z}{1+z} - \gamma\delta\right] \tag{13}$$

$$\alpha\gamma\delta f\left(i_L^h\right) + \left[\Delta_k + \overline{\theta}\left(1+i_B\right) f\left(i_L^h\right)\right]\left[\tfrac{\gamma+z}{1+z} - \gamma\delta\right] = \tfrac{(1-\gamma)z^2}{1+z}\xi_j + \frac{\Delta_k+\overline{\theta}(1+i_B)f\left(i_L^h\right)-2z\xi_j}{1+\frac{\partial\xi_j}{\partial\xi_k}} \tag{14}$$

$$\frac{\Delta_k+\overline{\theta}(1+i_B)f\left(i_L^h\right)-2z\xi_j}{(1+z)\xi_j}\,\frac{\frac{\partial\xi_j}{\partial i_L^h}}{1+\frac{\partial\xi_j}{\partial\xi_k}} = \frac{1-\pi}{1-\overline{\theta}}\left[\delta\left[\alpha-\overline{\theta}\left(1+i_B\right)\right] + \frac{\overline{\theta}(1+i_B)}{1+z} + \frac{\psi\left[\frac{2(1-\overline{\theta})}{1-\pi}f\left(i_L^h\right)-Y(\alpha)\right]}{\beta\xi_j^\gamma(1+z)^\gamma}\right] \tag{15}$$

Substitute in the reaction functions from the small bank's problem and rearrange. We need a triple $\left(z,\xi_j,i_L^h\right)$ that solves:

$$i_L^h = i_B + \frac{1}{\overline{\theta}(1-\pi)(1+i_B)}\left[\left(1+i_A\right)^2 - \left(1+i_B\right)^2 - \frac{\left(1-\overline{\theta}\right)\left[\frac{\alpha\gamma\delta\Delta_j}{\overline{\theta}(1+i_B)}+\left(1+\frac{\gamma+z}{1+z}\right)\xi_j\right]}{\frac{\gamma+z}{1+z}+\frac{\gamma\delta\Phi}{\overline{\theta}(1+i_B)}}\right]$$

$$i_L^h = i_B + \frac{(1+i_A)^2-(1+i_B)}{2(1-\pi)} - \frac{\delta\Phi}{2\psi} + \frac{\beta\overline{\theta}(1+i_B)\xi_j^\gamma(1+z)^\gamma}{2\psi(1-\gamma)}\left[1+\frac{\delta\Phi}{\overline{\theta}(1+i_B)} - \frac{\left[\Theta+2\left(z+\frac{\gamma+z}{1-\gamma}\right)\xi_j\right]\left[\frac{\gamma+z}{1+z}+\frac{\gamma\delta\Phi}{\overline{\theta}(1+i_B)}\right]^2}{\frac{\alpha\gamma\delta\Delta_j}{\overline{\theta}(1+i_B)}+\left[\frac{1+z}{\gamma+z}+\frac{\gamma+z}{1+z}+\frac{2\gamma\delta\Phi}{\overline{\theta}(1+i_B)}\right]\frac{(\gamma+z)\xi_j}{1-\gamma}}\right]$$

$$2q\left(z\right)\xi_j^2 - \left[\Theta m\left(z\right) + n\left(z\right) g\left(i_L^h\right)\right]\xi_j - \left(1-\gamma\right)\left[\gamma\Theta\left[1-\delta\left(1+z\right)\right] + zg\left(i_L^h\right)\right] g\left(i_L^h\right) = 0$$

To confirm the initial supposition that only small banks are constrained by the loan limit, we now need to check:

$$i_L^h \overset{?}{<} i_B + \frac{(1+i_A)^2-(1+i_B)}{1-\pi} \tag{16}$$

$$\lambda_k \equiv \frac{R_k}{D_k+W_k} = \pi\left(\theta_h-\theta_\ell\right)\left(1+i_B\right) + \frac{\overline{\theta}(1+i_B)-\psi\left(i_L^h-i_B\right)-\alpha\delta\left[1-\beta\xi_j^\gamma(1+z)^\gamma\right]}{1-\delta-\beta\xi_j^\gamma(1+z)^{\gamma-1}(1-\delta-\delta z)} \overset{?}{>} \alpha \tag{17}$$

This is what the gray area in Figure 17 is about. The red area focuses on comparative statics. In addition to i_L^h, we are interested in the big bank's liquidity ratio (λ_k), the total amount of credit (TL), and the fraction of credit extended off balance sheet (OBS). We can write these objects as:

$$\lambda_k = \pi\left(\theta_h-\theta_\ell\right)\left(1+i_B\right) + \frac{\overline{\theta}(1+i_B)-\psi\left(i_L^h-i_B\right)-\alpha\delta\left[1-\beta\xi_j^\gamma(1+z)^\gamma\right]}{1-\delta-\beta\xi_j^\gamma(1+z)^{\gamma-1}\left[1-\delta(1+z)\right]}$$

$$TL = 1 - \overline{\theta}\left(1+i_B\right) + \Psi\left(i_L^h\right) - \pi\left(\theta_h - \theta_\ell\right)\left(1+i_B\right)\left[1 - \delta - \beta\xi_j^\gamma\left(1+z\right)^{\gamma-1}\left[1 - \delta\left(1+z\right)\right]\right]$$

$$OBS = \frac{\beta\xi_j^\gamma(1+z)^{\gamma-1}}{TL}$$

To avoid carrying around heavy notation, define $X \equiv \frac{\partial\xi_j}{\partial i_L^h}$ and $Q \equiv 1 + \frac{\partial\xi_j}{\partial\xi_k}$. Also define $h\left(z\right) \equiv \frac{\gamma+z}{1+z} - \gamma\delta$ and $U \equiv \Delta_k + \overline{\theta}\left(1+i_B\right) f\left(i_L^h\right) - 2z\xi_j$. Differentiate equation (13) to get:

$$\underbrace{\left[1+\frac{\gamma+z}{1+z}\right]}_{\Omega_0}\frac{d\xi_j}{d\alpha} = \gamma\delta f\left(i_L^h\right) + \underbrace{\frac{(1-\gamma)\left[\Delta_j+\overline{\theta}(1+i_B)f\left(i_L^h\right)-\xi_j\right]}{(1+z)^2}}_{\Omega_1}\frac{dz}{d\alpha} - \underbrace{\frac{(1-\pi)\left[\alpha\gamma\delta+\overline{\theta}(1+i_B)h(z)\right]}{1-\overline{\theta}}}_{\Omega_2}\frac{di_L^h}{d\alpha}$$

Next, combine (13) and (14) then differentiate to get:

$$\underbrace{\left[2z + \frac{U}{Q}\frac{\partial Q}{\partial\xi_j} + \frac{U+\Theta h(z)Q}{\xi_j}\right]}_{\Omega_3}\frac{d\xi_j}{d\alpha} = \underbrace{\left[\frac{(1-\gamma)Q\left[\Theta+(1+z)^2\xi_j-2\xi_j\right]}{(1+z)^2} - 2\xi_j - \frac{U}{Q}\frac{\partial Q}{\partial z}\right]}_{\Omega_4}\frac{dz}{d\alpha} - \underbrace{\left[\frac{\overline{\theta}(1-\pi)(1+i_B)}{1-\overline{\theta}} + \frac{U}{Q}\frac{\partial Q}{\partial i_L^h}\right]}_{\Omega_5}\frac{di_L^h}{d\alpha}$$

$$where$$

$$\frac{\partial Q}{\partial\xi_j} = -\frac{1+z-Qz}{\xi_j}\left[Q - 1 + \frac{1-\gamma}{1+\gamma+2z}\left(1 - \frac{Qz}{1+z}\right)\right]$$

$$\frac{\partial Q}{\partial z} = -\left(Q-1\right)\left[Q - 1 + \frac{3+4z+\gamma}{1+\gamma+2z}\left(1 - \frac{Qz}{1+z}\right)\right]$$

$$\frac{\partial Q}{\partial i_L^h} = -\frac{\overline{\theta}(1-\pi)(1+i_B)}{1-\overline{\theta}}\frac{1+z-Qz}{\xi_j}\left(\frac{1-\gamma}{1+\gamma+2z}\right)\left(1 - \frac{Qz}{1+z}\right)$$

Finally, differentiate (15) to get:

$$\underbrace{\left[\frac{U}{Q}\frac{\partial X}{\partial\xi_j} + X\left[1 + \frac{\gamma+z}{1+z} - \frac{(1-\gamma)z^2}{1+z}\right] - \frac{\gamma(1-\pi)\left[\overline{\theta}(1+i_B)+\delta\Phi(1+z)\right]}{1-\overline{\theta}} - \frac{(1-\gamma)UX}{Q\xi_j}\right]}_{\Omega_6}\frac{d\xi_j}{d\alpha}$$

$$= \underbrace{\frac{\delta(1-\pi)}{1-\overline{\theta}}\left[1 - \frac{1}{\beta\xi_j^\gamma(1+z)^\gamma}\right]\left(1+z\right)\xi_j - \frac{U}{Q}\frac{\partial X}{\partial\alpha}}_{\Omega_7} - \underbrace{\left[\frac{U}{Q}\frac{\partial X}{\partial i_L^h} + \frac{2\psi(1-\pi)}{\beta\left(1-\overline{\theta}\right)}\xi_j^{1-\gamma}\left(1+z\right)^{1-\gamma}\right]}_{\Omega_8}\frac{di_L^h}{d\alpha}$$

$$-\underbrace{\left[\frac{U}{Q}\frac{\partial X}{\partial z} + \frac{1-\pi}{1-\overline{\theta}}\frac{\overline{\theta}(1+i_B)}{1+z}\xi_j - \frac{(1-\gamma)X\left[\Theta+(1+z)^2\xi_j-2\xi_j\right]}{(1+z)^2} - \frac{\gamma(1-\pi)}{1-\overline{\theta}}\left[\delta\Phi + \frac{\overline{\theta}(1+i_B)}{1+z}\right]\xi_j - \frac{(1-\gamma)UX}{Q(1+z)}\right]}_{\Omega_9}\frac{dz}{d\alpha}$$

$$where$$

$$\frac{\partial X}{\partial \alpha} = \frac{\gamma \delta X}{\alpha\gamma\delta + \overline{\theta}(1+i_B)h(z)}$$

$$\frac{\partial X}{\partial \xi_j} = \frac{X}{\xi_j}\left[1 + \frac{\left(1-\overline{\theta}\right)X}{1-\pi}\frac{2-\frac{(1-\gamma)(1+2z)}{(1+z)^2}}{\alpha\gamma\delta+\overline{\theta}(1+i_B)h(z)}\right]$$

$$\frac{\partial X}{\partial z} = \frac{X}{1+z}\left[1 - \frac{1}{z} + \frac{\overline{\theta}(1+i_B)(1-\gamma) - \frac{1-\overline{\theta}}{1-\pi}\frac{1+\gamma-2z^2}{z}X}{\left[\alpha\gamma\delta+\overline{\theta}(1+i_B)h(z)\right](1+z)}\right]$$

$$\frac{\partial X}{\partial i_L^h} = -\frac{X^2}{\alpha\gamma\delta+\overline{\theta}(1+i_B)h(z)}\frac{\overline{\theta}(1+i_B)(1-\gamma)z}{(1+z)^2\xi_j}$$

Combine the differentiated expressions to isolate the core derivatives:

$$\frac{di_L^h}{d\alpha} = \frac{\Omega_7\left[\Omega_1\Omega_3 - \Omega_0\Omega_4\right] + \gamma\delta f\left(i_L^h\right)\left[\Omega_3\Omega_9 + \Omega_4\Omega_6\right]}{\Omega_8\left[\Omega_1\Omega_3 - \Omega_0\Omega_4\right] + \Omega_2\left[\Omega_3\Omega_9 + \Omega_4\Omega_6\right] - \Omega_5\left[\Omega_0\Omega_9 + \Omega_1\Omega_6\right]}$$

$$\frac{dz}{d\alpha} = \frac{\Omega_0\Omega_7 - \gamma\delta f\left(i_L^h\right)\Omega_6}{\Omega_0\Omega_9 + \Omega_1\Omega_6} - \frac{\Omega_0\Omega_8 - \Omega_2\Omega_6}{\Omega_0\Omega_9 + \Omega_1\Omega_6}\frac{di_L^h}{d\alpha}$$

$$\frac{d\xi_j}{d\alpha} = \frac{\gamma\delta f\left(i_L^h\right)}{\Omega_0} + \frac{\Omega_1}{\Omega_0}\frac{dz}{d\alpha} - \frac{\Omega_2}{\Omega_0}\frac{di_L^h}{d\alpha}$$

We can then write the derivatives for the objects of interest as:

$$\begin{aligned}\frac{d\lambda_k}{d\alpha} \quad \overset{sign}{=} \quad & -\psi\frac{di_L^h}{d\alpha} - \delta\left[1 - \beta\xi_j^{\gamma}\left(1+z\right)^{\gamma}\right] + \alpha\beta\gamma\delta\xi_j^{\gamma}\left(1+z\right)^{\gamma}\left[\frac{1}{\xi_j}\frac{d\xi_j}{d\alpha} + \frac{1}{1+z}\frac{dz}{d\alpha}\right] \\ & + \frac{\beta\left[\lambda_k - \pi(\theta_h-\theta_\ell)(1+i_B)\right]\xi_j^{\gamma}}{(1+z)^{1-\gamma}}\left[\frac{\gamma\left[1-\delta(1+z)\right]}{\xi_j}\frac{d\xi_j}{d\alpha} - \left[\frac{1-\gamma}{1+z} + \gamma\delta\right]\frac{dz}{d\alpha}\right]\end{aligned}$$

$$\frac{dTL}{d\alpha} = \psi\frac{di_L^h}{d\alpha} + \frac{\beta\pi(\theta_h-\theta_\ell)(1+i_B)\xi_j^{\gamma}}{(1+z)^{1-\gamma}}\left[\frac{\gamma\left[1-\delta(1+z)\right]}{\xi_j}\frac{d\xi_j}{d\alpha} - \left[\frac{1-\gamma}{1+z} + \gamma\delta\right]\frac{dz}{d\alpha}\right]$$

$$\frac{dOBS}{d\alpha} \overset{sign}{=} \frac{\gamma}{\xi_j}\frac{d\xi_j}{d\alpha} - \frac{1-\gamma}{1+z}\frac{dz}{d\alpha} - \frac{1}{TL}\frac{dTL}{d\alpha}$$

Appendix C – Derivations for Section 4.5

The solvency condition for the placement bank is:

$$1 + \pi i_B + (1 - \pi) i_L^h + \varepsilon < (1 + i_A)^2 \tag{18}$$

We have written (18) under the assumption that the placement bank can lend all the funds it attracts (i.e., placements do not count as deposits so neither the loan-to-deposit ratio nor the reserve requirement restricts how much the placement bank can lend). Changing this assumption does not change the qualitative results.

Small Bank Problem Under (18), a bank will never hold placements that do not help it relax (7). That is, it will never choose x_j greater than $\overline{x}_j$, where $\overline{x}_j$ solves $\frac{\alpha - \overline{x}_j}{1 - \overline{x}_j} \equiv \alpha_0$. We can thus write the optimization problem of small bank j as:

$$\max_{\xi_j, \tau_j, R_j, x_j} \left\{ \begin{array}{l} \left[x_j \left[1 + \pi i_B + (1 - \pi) i_L^h + \varepsilon\right] + (1 - x_j)(1 + i_A)^2\right] \left[D_j(\xi_j) + W_j(\xi_j) - R_j\right] \\ + \left[1 + \pi i_B + (1 - \pi) i_L^h\right] \left[R_j - \overline{\theta}(1 + i_B)\left[D_j(\xi_j) + W_j(\xi_j)\right]\right] \\ - \left(1 - \overline{\theta}\right) \left[(1 + i_B)^2 \left[D_j(\xi_j) + W_j(\xi_j)\right] + \xi_j W_j(\xi_j)\right] - c\tau_j \end{array} \right\}$$

$$\textit{subject to}$$

$$R_j \geq \tfrac{\alpha - x_j}{1 - x_j} \left[D_j(\xi_j) + (1 - \tau_j) W_j(\xi_j)\right]$$

$$x_j \in \left[0, \tfrac{\alpha - \alpha_0}{1 - \alpha_0}\right]$$

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

As before, the multiplier on the reserve constraint is denoted by μ_j, the multiplier on $\tau_j \geq 0$ is denoted by η_j^0, and the multiplier on $\tau_j \leq 1$ is denoted by η_j^1. Also let $\widetilde{\eta}_j^0$ and $\widetilde{\eta}_j^1$ denote the multipliers on $x_j \geq 0$ and $x_j \leq \frac{\alpha - \alpha_0}{1 - \alpha_0}$ respectively. We again define the constant Δ_j and the function $f\left(i_L^h\right)$.[36] The first order conditions with respect to R_j, τ_j, x_j, and ξ_j are:

[36]Remember $\Delta_j \equiv \frac{\left[1 - \overline{\theta}(1 + i_B)\right](1 + i_A)^2}{1 - \overline{\theta}} - (1 + i_B)^2$ and $f\left(i_L^h\right) \equiv \frac{(1 + i_A)^2 - \left[1 + \pi i_B + (1 - \pi) i_L^h\right]}{1 - \overline{\theta}}$.

$$\mu_j = \left(1-\overline{\theta}\right)\left(1-x_j\right) f\left(i_L^h\right) + x_j \varepsilon \tag{19}$$

$$\eta_j^1 = \eta_j^0 + \mu_j \tfrac{\alpha - x_j}{1-x_j} W_j\left(\xi_j\right) - c \tag{20}$$

$$\widetilde{\eta}_j^1 = \widetilde{\eta}_j^0 + \mu_j \tfrac{1-\alpha}{(1-x_j)^2}\left[D_j\left(\xi_j\right) + \left(1-\tau_j\right) W_j\left(\xi_j\right)\right] - \tfrac{\mu_j - \varepsilon}{1-x_j}\left[D_j\left(\xi_j\right) + W_j\left(\xi_j\right) - R_j\right] \tag{21}$$

$$\xi_j + \tfrac{W_j\left(\xi_j\right)}{W_j'\left(\xi_j\right)} = \left[\Delta_j - \tfrac{\alpha - \overline{\theta}(1+i_B)}{1-\overline{\theta}} \tfrac{\mu_j}{1-x_j} + \tfrac{1-\overline{\theta}(1+i_B)}{1-\overline{\theta}} \tfrac{x_j \varepsilon}{1-x_j}\right]\left[1 + \tfrac{D_j'\left(\xi_j\right)}{W_j'\left(\xi_j\right)}\right] + \tfrac{\mu_j}{1-x_j} \tfrac{\alpha - x_j}{1-\overline{\theta}} \tau_j \tag{22}$$

Recall that we focus on $\mu_j > 0$ so the constraint on R_j binds for small banks. We can then substitute the expression for R_j into (21) to get:

$$\widetilde{\eta}_j^1 = \widetilde{\eta}_j^0 + \left[\tfrac{1-\alpha}{1-x_j}\left[D_j\left(\xi_j\right) + W_j\left(\xi_j\right)\right] + \tfrac{\alpha - x_j}{1-x_j} \tau_j W_j\left(\xi_j\right)\right] \tfrac{\varepsilon}{1-x_j} - \tfrac{\mu_j}{1-x_j} \tau_j W_j\left(\xi_j\right) \tag{23}$$

No Crackdown on Trusts This case amounts to $c = 0$. Equation (20) implies $\eta_j^1 > 0$ and thus $\tau_j = 1$. To then get $x_j = 0$ from equations (19) and (23), we need:

$$f\left(i_L^h\right) > \left[1 + (1-\alpha) \tfrac{D_j\left(\xi_j\right)}{W_j\left(\xi_j\right)}\right] \tfrac{\varepsilon}{1-\overline{\theta}} \tag{24}$$

As long as condition (24) is true, equation (22) gives the same expression for ξ_j as the model without placements, namely equation (12) in Appendix B.

Full Crackdown on Trusts We interpret this case using $c \to \infty$. Equation (20) implies $\eta_j^0 > 0$ and thus $\tau_j = 0$. If $\varepsilon > 0$, then equation (23) yields $\widetilde{\eta}_j^1 > 0$ and thus $x_j = \frac{\alpha - \alpha_0}{1-\alpha_0}$. If $\varepsilon = 0$, then equation (23) yields $\widetilde{\eta}_j^1 = \widetilde{\eta}_j^0 = 0$ which means the small bank is indifferent over any $x_j \in \left[0, \frac{\alpha - \alpha_0}{1-\alpha_0}\right]$. For the sake of argument, let's suppose it still chooses $x_j = \frac{\alpha - \alpha_0}{1-\alpha_0}$ when $\varepsilon = 0$. The combination of equations (19) and (22) then yields ξ_j implicitly defined by:

$$\left[1+\frac{\xi_k+\gamma\xi_j}{\xi_j+\xi_k}\right]\xi_j=\left[\Delta_j-\left[\alpha-\overline{\theta}\left(1+i_B\right)\right]f\left(i_L^h\right)+\frac{\alpha-\alpha_0}{1-\overline{\theta}}\varepsilon\right]\left[\frac{\xi_k+\gamma\xi_j}{\xi_j+\xi_k}-\delta\gamma\right] \tag{25}$$

Differentiating equation (25) yields the following reaction functions:

$$\frac{d\xi_j}{d\xi_k}=\frac{\xi_j}{\xi_k+\frac{\left(\xi_j+\xi_k\right)\left[(1+\gamma)\xi_j+2\xi_k\right]}{(1-\gamma)\left[\Delta_j-\left[\alpha-\overline{\theta}(1+i_B)\right]f\left(i_L^h\right)+\frac{\alpha-\alpha_0}{1-\overline{\theta}}\varepsilon-\xi_j\right]}}$$

$$\frac{d\xi_j}{di_L^h}=\frac{\frac{1-\pi}{1-\overline{\theta}}\left[\alpha-\overline{\theta}(1+i_B)\right]\left[1-\frac{\delta\gamma\left(\xi_j+\xi_k\right)}{\xi_k+\gamma\xi_j}\right]}{1+\frac{\xi_j+\xi_k}{\xi_k+\gamma\xi_j}+\left[\Delta_j-\left[\alpha-\overline{\theta}(1+i_B)\right]f\left(i_L^h\right)+\frac{\alpha-\alpha_0}{1-\overline{\theta}}\varepsilon-\xi_j\right]\frac{(1-\gamma)\xi_k}{\xi_j+\xi_k}\frac{1}{\xi_k+\gamma\xi_j}}$$

Big Bank Problem As before, we focus on parameters such that the big bank is not constrained by regulation. There is then no reason for the big bank to invest in placements since $(1+i_A)^2$ dominates the return from placements. Mathematically, $\tau_k=x_k=0$ and the structure of the big bank's problem is as in the model without placements. If $c=0$ and condition (24) is satisfied, then the problem is exactly the same. If $c\to\infty$, then the big bank still chooses ξ_k and i_L^h to maximize:

$$\begin{aligned}&(1+i_A)^2\left(D_k+W_k-R_k\right)+\left[1+\pi i_B+(1-\pi)\,i_L^h\right]\left[R_k-\overline{\theta}\left(1+i_B\right)\left(D_k+W_k\right)\right]\\&-\left(1-\overline{\theta}\right)\left[(1+i_B)^2\left(D_k+W_k\right)+\xi_kW_k\right]-\pi\left(1-\pi\right)\left(\theta_h-\theta_\ell\right)\left(1+i_B\right)\left(i_L^h-i_B\right)\left(D_k+W_k\right)\end{aligned}$$

But it is now subject to:

$$R_k=\overline{\theta}\left(1+i_B\right)+\pi\left(\theta_h-\theta_\ell\right)\left(1+i_B\right)\left(D_k+W_k\right)-\Psi\left(i_L^h\right)-\frac{\alpha-x_j}{1-x_j}\left[D_j+\left(1-\tau_j\right)W_j\right]$$

And equation (25) with $\tau_j=0$ and $x_j=\frac{\alpha-\alpha_0}{1-\alpha_0}$. First order condition with respect to ξ_k is:

$$\Delta_k+\left[\overline{\theta}\left(1+i_B\right)-\alpha_0\right]f\left(i_L^h\right)=\frac{\left[2\xi_j+(1+\gamma)\xi_k-(1-\gamma)\xi_k\frac{\partial\xi_j}{\partial\xi_k}\right]\xi_k}{[1-\gamma(1-\delta)]\xi_j+\gamma\delta\xi_k-\left[(1-\gamma\delta)\xi_k+\gamma(1-\delta)\xi_j\right]\frac{\partial\xi_j}{\partial\xi_k}}$$

The first order condition with respect to i_L^h is:

$$= \frac{\Delta_k + \left[\overline{\theta}\,(1+i_B) - \alpha_0\right] f\left(i_L^h\right)}{\frac{(1-\gamma)\xi_k^2 \frac{\partial \xi_j}{\partial i_L^h} + \frac{\psi\left[f\left(i_L^h\right) - \frac{(1-\pi)\left(i_L - i_B\right)}{1-\overline{\theta}}\right]}{\beta\left(\xi_j+\xi_k\right)^{\gamma-2}} + \frac{(1-\pi)\left[\overline{\theta}(1+i_B)-\alpha_0\right]}{1-\overline{\theta}}\left[\frac{\delta}{\beta\left(\xi_j+\xi_k\right)^{\gamma-2}} + \left(\xi_j+\xi_k\right)\left[\xi_j - \delta\left(\xi_j+\xi_k\right)\right]\right]}{\left[(1-\gamma\delta)\xi_k + \gamma(1-\delta)\xi_j\right]\frac{\partial \xi_j}{\partial i_L^h}}}$$

Summary of Equilibrium Conditions for $c \to \infty$ Recall $z \equiv \xi_k/\xi_j$. Putting all the equations together and rearranging yields:

$$\xi_j = \left[\Delta_j + \frac{\alpha-\alpha_0}{1-\overline{\theta}}\varepsilon + \left[\overline{\theta}\,(1+i_B) - \alpha\right] f\left(i_L^h\right)\right] \frac{\gamma+z-\delta\gamma(1+z)}{1+\gamma+2z}$$

$$\frac{1+\gamma+2z}{\gamma+z-\delta\gamma(1+z)} \frac{\frac{\gamma+z}{1+z}\left[2(\gamma+z)+\frac{(1-\gamma)(1+\gamma+2z)}{\gamma+z-\delta\gamma(1+z)}\right]-(1+\gamma\delta)\left[2(\gamma+z)+\frac{(1-\gamma)(\gamma+2z)}{\gamma+z-\delta\gamma(1+z)}\right]}{\frac{(1-\gamma)z^2}{1+z}\left[2(\gamma+z)+\frac{(1-\gamma)(1+\gamma+2z)}{\gamma+z-\delta\gamma(1+z)}\right]-2z\left[2(\gamma+z)+\frac{(1-\gamma)(\gamma+2z-\delta\gamma)}{\gamma+z-\delta\gamma(1+z)}\right]} = \frac{\Delta_j+\frac{\alpha-\alpha_0}{1-\overline{\theta}}\varepsilon+\left[\overline{\theta}(1+i_B)-\alpha\right]f\left(i_L^h\right)}{\Delta_k+\left[\overline{\theta}(1+i_B)-\alpha_0\right]f\left(i_L^h\right)}$$

$$f\left(i_L^h\right) = \frac{\frac{\left[\overline{\theta}(1+i_B)-\alpha\right]\left[\frac{\gamma+z}{1+z}-\delta\gamma\right]}{2(\gamma+z)+\frac{(1-\gamma)(1+\gamma+2z)}{\gamma+z-\delta\gamma(1+z)}}\left[2z-\frac{\Delta_k}{\xi_j}\right]+\left[\overline{\theta}(1+i_B)-\alpha_0\right]\left[\delta-\frac{1}{1+z}\right]-\frac{\delta(\alpha-\alpha_0)-\psi Y(\alpha)}{\beta\xi_j^{\gamma}(1+z)^{\gamma}}}{\frac{2\psi\left(1-\overline{\theta}\right)}{\beta(1-\pi)\xi_j^{\gamma}(1+z)^{\gamma}}+\frac{\left[\overline{\theta}(1+i_B)-\alpha\right]\left[\overline{\theta}(1+i_B)-\alpha_0\right]\left[\frac{\gamma+z}{1+z}-\delta\gamma\right]}{\left[2(\gamma+z)+\frac{(1-\gamma)(1+\gamma+2z)}{\gamma+z-\delta\gamma(1+z)}\right]\xi_j}}$$

We then need to verify solvency of the placement bank, namely condition (18). We also need to verify that the constraint on R_j binds while that on R_k does not. In particular:

$$f\left(i_L^h\right) + \frac{\alpha-\alpha_0}{1-\alpha}\frac{\varepsilon}{1-\overline{\theta}} \overset{?}{\geq} 0$$

$$\frac{\overline{\theta}(1+i_B)-\alpha_0-\psi\left(i_L^h-i_B\right)}{1-\delta-\beta\xi_j^{\gamma}(1+z)^{\gamma}\left[\frac{1}{1+z}-\delta\right]} \overset{?}{\geq} \alpha-\alpha_0-\pi\left(\theta_h-\theta_\ell\right)(1+i_B)$$

Results are summarized in Figure 18.