BANK OF ENGLAND

Working Paper No. 529

Banks are not intermediaries of loanable funds — and why this matters

Zoltan Jakab and Michael Kumhof

May 2015

BANK OF ENGLAND

Working Paper No. 529

Banks are not intermediaries of loanable funds — and why this matters

Zoltan Jakab[(1)] and Michael Kumhof[(2)]

Abstract

In the intermediation of loanable funds model of banking, banks accept deposits of pre-existing real resources from savers and then lend them to borrowers. In the real world, banks provide financing through money creation. That is they create deposits of new money through lending, and in doing so are mainly constrained by profitability and solvency considerations. This paper contrasts simple intermediation and financing models of banking. Compared to otherwise identical intermediation models, and following identical shocks, financing models predict changes in bank lending that are far larger, happen much faster, and have much greater effects on the real economy.

Key words: Banks, financial intermediation, loanable funds, money creation, loans, deposits, leverage, spreads.

JEL classification: E44, E52, G21.

(1) International Monetary Fund. Email: zjakab@imf.org
(2) Bank of England. Email: michael.kumhof@bankofengland.co.uk

The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or the International Monetary Fund. This paper was finalised on 20 March 2015.

The Bank of England's working paper series is externally refereed.

Information on the Bank's working paper series can be found at
www.bankofengland.co.uk/research/Pages/workingpapers/default.aspx

Publications Team, Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 7601 4030 Fax +44 (0)20 7601 3298 email publications@bankofengland.co.uk

ISSN 1749-9135 (on-line)

Summary

Since the Great Recession, banks have increasingly been incorporated into macroeconomic models. However, this literature confronts many unresolved issues. This paper shows that many of them are attributable to the use of the intermediation of loanable funds (ILF) model of banking. In the ILF model, bank loans represent the intermediation of real *savings*, or loanable funds, between non-bank savers and non-bank borrowers. But in the real world, the key function of banks is the provision of *financing*, or the creation of new monetary purchasing power through loans, for a single agent that is both borrower and depositor. The bank therefore creates its own funding, deposits, in the act of lending, in a transaction that involves no intermediation whatsoever. Third parties are only involved in that the borrower/depositor needs to be sure that others will accept his new deposit in payment for goods, services or assets. This is never in question, because bank deposits are any modern economy's dominant medium of exchange.

Furthermore, if the loan is for physical investment purposes, this new lending and money is what triggers investment and therefore, by the national accounts identity of saving and investment (for closed economies), saving. Saving is therefore a consequence, not a cause, of such lending. Saving does not finance investment, financing does. To argue otherwise confuses the respective macroeconomic roles of resources (saving) and debt-based money (financing).

The paper shows that this financing through money creation (FMC) description of the role of banks can be found in many publications of the world's leading central banks. What has been much more challenging is the incorporation of the FMC view's insights into dynamic stochastic general equilibrium (DSGE) models that can be used to study the role of banks in macroeconomic cycles. DSGE models are the workhorse of modern macroeconomics, and are a key tool in macro-prudential policy analysis. They study the interactions of multiple economic agents that optimise their utility or profit objectives over time, subject to budget constraints and random shocks.

The key contribution of this paper is therefore the development of the essential ingredients of DSGE models with FMC banks, and a comparison of their predictions with those of otherwise identical DSGE models with ILF banks. The result of our model comparison exercise is that, compared to ILF models, and following identical shocks to financial conditions that affect the creditworthiness of bank borrowers, FMC models predict changes in the size of bank balance sheets that are far larger, happen much faster, and have much greater effects on the real economy, while the adjustment process depends far less on changes in lending spreads, the dominant adjustment channel in ILF models. Compared to ILF models, FMC models also predict pro-cyclical rather than countercyclical bank leverage, and a significant role for quantity rationing of credit rather than price rationing during downturns. We show that these predictions of FMC models are much more in line with the stylised facts than those of ILF models.

The fundamental reason for these differences is that savings in the ILF model of banking need to be accumulated through a process of either producing additional goods or foregoing consumption of existing goods, a physical process that by its very nature is *slow* and *continuous*. On the other hand, FMC banks that create purchasing power can technically do so *instantaneously* and *discontinuously*, because the process does not involve physical goods, but rather the creation of money through the simultaneous expansion of both sides of banks' balance sheets. While money is essential to facilitating purchases and sales of real resources outside the banking system, it is not itself a physical resource, and can be created

at near zero cost. In other words, the ILF model is fundamentally a model of banks as barter institutions, while the FMC model is fundamentally a model of banks as monetary institutions.

The fact that banks *technically* face no limits to increasing the stocks of loans and deposits instantaneously and discontinuously does not, of course, mean that they do not face other limits to doing so. But the most important limit, especially during the boom periods of financial cycles when all banks simultaneously decide to lend more, is their own assessment of the implications of new lending for their profitability and solvency, rather than external constraints such as loanable funds, or the availability of central bank reserves.

This finally takes us to the venerable deposit multiplier (DM) model of banking, which suggests that the availability of central bank high-powered money imposes another limit to rapid changes in the size of bank balance sheets. The DM model however does not recognise that modern central banks target interest rates, and are committed to supplying as many reserves (and cash) as banks demand at that rate. The quantity of reserves is therefore a consequence, not a cause, of lending and money creation.

Contents

Tables

Figures

I. Introduction

In the wake of the 2007/8 financial crisis, the role of banks in the economy has attracted more attention than at any time since the 1930s, with policymakers clearly recognising the importance of a healthy banking system for the real economy. Macroeconomic theory was however initially not ready to provide much support in studying the interactions between banks and the real economy, as banks were not a part of most macroeconomic models. The reason is that for many decades the private banking system had not been seen as an important source of vulnerability, so that almost all interest in banks and in prudential banking regulation was of a microeconomic nature. This is in stark contrast to the preoccupation of the leading macroeconomists of the 1920s, 1930s and 1940s with the problems of banking[1], which after the 1950s continued only in a small part of the profession, with the work of some post-Keynesians.[2]

The Great Recession changed this dramatically. Among policymakers, this culminated in the recent debates over the Basel III framework and other regulatory initiatives. Academic macroeconomics also started to pay attention to the role of banks and of prudential banking regulation. However, as emphasised by Adrian, Colla and Shin (2013), in this new literature there are many unresolved issues. We will show in this paper that many of these are due to the fact that this literature is almost without exception based on a version of the intermediation of loanable funds (ILF) model of banking.[3]

In the simple ILF model, bank loans represent the intermediation of real *savings*, or loanable funds, from non-bank savers to non-bank borrowers. Lending starts with banks collecting deposits of real savings from one agent, and ends with the lending of those savings to another agent. In the real world, the key function of banks is the provision of *financing*, or the creation of new monetary purchasing power through loans, for a single agent that is both borrower and depositor.[4] Specifically, whenever a bank makes a new loan to a non-bank customer X, it creates a new loan entry in the name of customer X on the asset side of its balance sheet, and it simultaneously creates a new and equal-sized deposit entry, also in the name of customer X, on the liability side of its balance sheet. The bank therefore creates its own funding, deposits, in the act of lending. And because both entries are in the name of customer X, there is no intermediation whatsoever at the moment when a new loan is made. No real resources need to be diverted from other uses, by other agents, in order to be able to lend to customer X. What is needed from third parties is only the acceptance of the newly created purchasing power in payment for goods and services. This is never in question, because bank demand deposits are any modern economy's dominant medium of exchange, in other words its money.[5]

Furthermore, if the loan is for physical investment purposes, this new lending and money is what triggers investment and therefore, by the national accounts identity of saving and investment (for

[1] Examples include Knight (1927, 1933), Douglas (1935), Fisher (1935, 1936), Graham (1936), Simons (1946, 1948) and Schumpeter (1954).

[2] Key references include Moore (1979, 1983), Graziani (1989) and Minsky (1986, 1991).

[3] In undergraduate textbooks one also finds the older deposit multiplier (DM) model of banking, but this has not featured at all in the recent academic literature. We will nevertheless discuss it later in this paper, because of its enduring influence on popular understandings of banking.

[4] The key distinction between saving and financing has for some time been emphasised by researchers at the BIS (see, for example, Borio and Disyatat (2011)).

[5] Bank deposits can fulfill this role because the central bank and/or government, though a combination of deposit insurance, prudential regulation and lender of last resort functions, ensures that bank deposits are considered safe by the public, and therefore trade at par with base money. See McLeay, Radia and Thomas (2014a,b).

closed economies), saving. Saving is therefore a consequence, not a cause, of such lending. Saving does not finance investment, financing does.[6] To argue otherwise confuses the respective macroeconomic roles of resources (saving) and debt-based money (financing).

This financing through money creation (FMC) description of the role of banks can be found in many publications of the world's leading central banks. See McLeay, Radia and Thomas (2014a,b) for an excellent summary, and Section II of this paper for a much more comprehensive literature survey and exposition. What has been much more challenging is the incorporation of the insights from the FMC view into macroeconomic models that can be used to understand the role of banks in macroeconomic cycles.[7]

The key contribution of this paper is therefore the development of the essential ingredients of dynamic stochastic general equilibrium (DSGE) models with FMC banks, and a comparison of their predictions with those of otherwise identical DSGE models with ILF banks. After this we will also show, in Section VI, that the predictions for key variables with FMC models are much more in line with the stylised facts than with ILF models. In the literature such stylised facts are typically presented first, followed by the model. This is because the key question is usually whether any new model can be motivated by the evidence, and shown to be more consistent with the evidence than competing models. In the present context this would amount to asking whether we can provide empirical evidence for the "theory" that banks create money through loans, rather than intermediating pre-existing savings. But this is not a theory that needs to be proved, it is a simple fact, it is part of the elementary design of any modern economy's financial system. The empirical evidence in Section VI is therefore not critical for justifying our modeling of banks. But it is critical for demonstrating that these insights have quantitatively important consequences. It is for this reason alone that we study the stylised facts.

The result of our model comparison exercise is that, compared to ILF models, and following identical shocks, FMC models predict changes in the size of bank balance sheets that are far larger, happen much faster, and have much greater effects on the real economy, while the adjustment process depends far less on changes in lending spreads. Compared to ILF models, FMC models also predict procyclical rather than countercyclical bank leverage, and an important role for quantity rationing of credit, rather than an almost exclusive reliance on price rationing, in response to contractionary financial shocks.

The fundamental reason for these differences is that savings in the ILF model of banking need to be accumulated through a process of either producing additional goods or foregoing consumption of existing goods, a physical process that by its very nature is *slow* and *continuous*. On the other hand, FMC banks that create purchasing power can technically do so *instantaneously* and *discontinuously*[8], because the process does not involve (physical) goods, but rather the creation of

[6] This result is very general, it applies to all investment, not only to investment financed through bank loans, see Lindner (2012, 2013). *Financial* saving is a zero-sum game, as aggregate financial saving cannot increase through individual financial saving decisions, only through additional financing, typically loans. On the other hand, in a closed economy, *macroeconomic* (national accounts) saving is equal to investment by accounting definition rather than as a result of equilibrium, and the quantity of that saving is unrelated to the overall quantity of financing.

[7] FMC models can be used to highlight the dangers of excessive credit expansions. But not all credit expansions are excessive, and FMC models also highlight that, following an improvement in economic fundamentals, banks can greatly enhance an economy's ability to reach a higher level of output, by flexibly providing the economy with the necessary purchasing power.

[8] We use the term discontinuous by analogy with continuous time models. The idea, as will become clearer below, is that jumps in credit and money can be so large because credit and money are not predetermined variables.

(digital) money through the simultaneous expansion of both sides of banks' balance sheets.[9] While money is essential to facilitating purchases and sales of real resources outside the banking system, it is not itself a physical resource, and can be created at near zero cost. In different words, and as shown in more detail in Section II, the ILF model is fundamentally a model of banks as barter institutions, while the FMC model is fundamentally a model of banks as monetary institutions.

There is yet another way of stating this in terms of balance sheets. The ILF model looks at banks as institutions that record *nonzero net* non-financial (goods) transactions, which by their nature require saving of real resources to take place before any lending. The FMC model looks at banks as institutions that record *nonzero gross, but zero net*, financial (money) transactions, which clearly do not require prior saving of real resources, but which are nevertheless essential for the functioning of the economy because the bank liability side of this transaction creates the economy's medium of exchange. This, the creation of gross positions with zero net principal value, but of course with a positive net interest flow to the bank over time, is precisely the meaning of bank financing, the very rationale for the existence of banks.

The fact that banks *technically* face no limits to increasing the stocks of loans and deposits instantaneously and discontinuously does not, of course, mean that they do not face other limits to doing so. But the most important limit, especially during the boom periods of financial cycles when all banks simultaneously decide to lend more, is their own assessment of the implications of new lending for their profitability and solvency. McLeay et al. (2014b) also make this point. They add that, from the (microeconomic) point of view of an individual bank that considers whether to deviate significantly from the behaviour of its competitors, other important limits exist, especially increased credit risk when lending too fast to marginal borrowers, and increased liquidity risk when creating deposits so fast that too many of them are lost to competitors.

The deposit multiplier (DM) model of banking suggests that the availability of central bank high-powered money (reserves or cash) imposes another limit to rapid changes in the size of bank balance sheets. In the deposit multiplier model, the creation of additional broad monetary aggregates requires a prior injection of high-powered money, because private banks can only create such aggregates by repeated re-lending of the initial injection. This view is fundamentally mistaken. First, it ignores the fact that central bank reserves cannot be lent to non-banks (and that cash is never lent directly but only withdrawn against deposits that have first been created through lending). Second, and more importantly, it does not recognise that modern central banks target interest rates, and are committed to supplying as many reserves (and cash) as banks demand at that rate, in order to safeguard financial stability.[10] The quantity of reserves is therefore a consequence, not a cause, of lending and money creation. This view concerning central bank reserves, like the FMC view of banks, has been repeatedly described in publications of the world's leading central banks.

The rest of the paper is organised as follows. Section II provides more detailed explanations concerning the misconceptions about banks and money in the ILF and DM models of banking, and contrasts these with the FMC model. Section III briefly reviews the existing theoretical

[9] The only tool that the bank requires to complete this process is a keyboard or, in earlier times, a pen. A particularly concise statement of this fact can be found in Friedman (1971, p. 2): "The correct answer for [the question of the origin of] both Euro-dollars and liabilities of U.S. banks is that their major source is a bookkeeper's pen."

[10] As shown by Kydland and Prescott (1990), the availability of central bank reserves did not even constrain banks during the period when the Fed officially targeted monetary aggregates.

literature on banks and relates it to the models studied in this paper. Section IV develops the theoretical models that will be used to study the differences between the ILF and FMC models of banking. Section V uses these models to generate and discuss illustrative simulations. Section VI presents the stylised facts as they relate to the predictions of the models. Section VII concludes.

II. Misconceptions about Banks in ILF and DM Models

Subsection A contrasts the ILF and FMC views of banking. We first cite authoritative statements that express the FMC view, including recent publications of the world's leading central banks, and leading economists of the past. Thereafter, using balance sheets, we study the problems with the ILF view in greater detail, and then explain that these problems can be corrected by adopting the FMC view. Subsection B discusses problems with the DM view of banking, again by citing leading central banks and economists. Subsection C adds brief comments on the roles of monetary and macroprudential policies.

A. ILF Models? New Loans Lead to Deposit Creation, Not Vice Versa

1. Statements by Central Banks and Early 20th Century Economists

The fact that banks create their own funds through lending is acknowledged in descriptions of the money creation process by leading central banks and policymaking authorities. The oldest goes back to Graham Towers (1939), the then governor of the central bank of Canada: "Each and every time a bank makes a loan, new bank credit is created – new deposits – brand new money". Berry, Harrison, Thomas and de Weymarn (2007), staff at the Bank of England: "When banks make loans, they create additional deposits for those that have borrowed the money." Keister and McAndrews (2009), staff economists at the Federal Reserve Bank of New York: "Suppose that Bank A gives a new loan of $20 to Firm X ... Bank A does this by crediting Firm X's account by $20. The bank now has a new asset (the loan to Firm X) and an offsetting liability (... Firm X's deposit at the bank)." Bundesbank (2012) (translation by the authors): "How is deposit money created? The procedure is equivalent to the creation of central bank money: As a rule the commercial bank extends a loan to a customer and credits the corresponding amount to his deposit account. ... The creation of deposit money is therefore an accounting transaction." Mervyn King (2012), former Governor of the Bank of England: "When banks extend loans to their customers, they create money by crediting their customers' accounts." Lord Adair Turner (2013), former head of the UK Financial Services Authority: "Banks do not, as many textbooks still suggest, take deposits of existing money from savers and lend it out to borrowers: they create credit and money ex nihilo – extending a loan to the borrower and simultaneously crediting the borrower's money account."[11] One can find similar statements from the private sector. One example is Standard and Poor's (2013): "Banks lend by simultaneously creating a loan asset and a deposit liability on their balance sheet. That is why it is called credit "creation" – credit is created literally out of thin air (or with the stroke of a keyboard)."

[11] Pozsar (2014), who provides a very detailed description of the institutional details of today's financial system, also emphasises that banks create money ex nihilo.

The fact that banks create their own funds through lending is also repeatedly emphasised in the older economics literature. One of the earliest statements is due to Wicksell (1906): “The lending operations of the bank will consist rather in its entering in its books a fictitious deposit equal to the amount of the loan...”. Rogers (1929): “... a large proportion of ... [deposits] under certain circumstances may be manufactured out of whole cloth by the banking institutions themselves.” The following passage from Schumpeter (1954) is highly illuminating (emphasis added): “But this ... makes it highly inadvisable to construe bank credit on the model of existing funds' being withdrawn from previous uses by an *entirely imaginary act of saving* and then lent out by their owners. It is much more realistic to say that the banks ... *create deposits in their act of lending*, than to say that they lend the deposits that have been entrusted to them. ... The theory to which economists clung so tenaciously makes [depositors] out to be savers when they neither save nor intend to do so; it attributes to them an influence on the "supply of credit" which they do not have. Nevertheless, it proved extraordinarily difficult for economists to recognise that bank loans and bank investments do create deposits. In fact, throughout the period under review they refused with practical unanimity to do so. And even in *1930, when a large majority had been converted* and accepted that doctrine as a matter of course, Keynes rightly felt it to be necessary to re-expound and to defend the doctrine at length ...”. The first half of this statement is exactly the FMC view. The second half shows that a struggle to convince the economics profession, and policymakers, of the FMC view had been won by 1930. This is also reflected in the report of the Macmillan Committee (1931).

Unfortunately, the work of Gurley and Shaw (1955, 1956) brought a major step backwards in our understanding of banks and money. Gurley and Shaw replaced the critical distinction between banks, which can create their own funds in the act of lending, and non-bank financial intermediaries, which cannot, with the far less important distinction between intermediated and direct debt. They treated banks as simply another form of intermediary, and bank liabilities as simply another form of debt. This work was heavily (and correctly) criticised by monetary theorists of that time, including Culbertson (1958, p. 121), who writes: “A change in the volume of demand deposits, in contrast, is initiated by banks when they change the volume of their debt holdings; the banks' creditors, as such, play no active role in the process. The banking system "creates credit" by acquiring debt and creating demand deposits to pay for it. The commercial banks do not need "to borrow loanable funds from spending units with surpluses" [as claimed by Gurley and Shaw] in order to extend credit...”. Similarly, Smith (1959) writes: “Commercial bank credit creation makes funds available to finance expenditures in excess of the funds arising out of the current income flow. ... Commercial banks ... are distinctly *not* intermediaries. That is, the decision to save a portion of current income and to hold the savings in the form of a demand deposit does not make any more funds available to the capital market than would have been available had the decision been made to spend instead, and does no more than to restore to the commercial banking *system* the lending power that was lost when the original cheque was written to transmit income to the recipient.”

Tobin (1963) played a critical role in establishing the ILF view of Gurley and Shaw as the new dominant paradigm. This paper stated explicitly that banks are not creators of money in the sense of the FMC view. Tobin's argument is that the behaviour of the agents that receive the newly created bank deposits after they are spent will be a function of their portfolio preferences and the endogenous adjustments of returns on deposits and alternative assets, with some agents using the new money to repay outstanding loans, thereby quickly destroying the money. In other words, banks do not possess the same “widow's cruse” as the central bank with its printing press, and money created by banks is not a "hot potato" that can be passed along by non-banks but whose aggregate quantity cannot be changed by them. This however is not a counterargument to the FMC view, because the FMC view does not make these claims. In fact, its claim is precisely that the extent of credit and money creation is determined by the interaction of the optimisation problems of banks and their customers, and that the solution to these problems is interior, in other words that the extent of credit and money creation is finite. It is simply not useful to frame this argument in a black-and-white fashion, whereby either banks do or do not possess the power to independently create additional credit and money, as some opponents of the FMC view have done using the Tobin (1963) paper.[12] Because, in order to challenge the FMC view in this fashion, one would have to argue that in general 100%, or close to 100%, of newly created bank money will be extinguished in the above-mentioned way in the short to medium run, so that money creation by banks cannot cause significant financial and real cycles. That however would be a very strange argument, of a kind that is never invoked for any other shocks, for example shocks to consumption demand. To see this, assume a credit supply shock whereby one group of agents receives larger loans and therefore larger money balances. This does not imply that another group of agents must automatically want to repay existing loans after receiving the additional money, just like a shock that increases the consumption demand of one group of agents does not imply that another group of agents must automatically want to reduce their consumption. In fact, in both cases, it implies the exact opposite. In the case of the credit supply shock, the reason is that the additional money creation stimulates additional economic activity, by facilitating additional transactions, which in turn means that households want to keep some of the additional money to support additional spending, rather than to repay existing loans. This phenomenon is very prominent in the simulations of our model. And it does not, contrary to what is alleged in Tobin (1963), depend on the assumption that banks create a special but hard-to-define liability called *money*. The critical insight is that banks can create their own *funds* instantaneously, and that there is a well-defined demand for those funds, whether they are called money or not. A portfolio-balance-type demand would be perfectly sufficient to generate similar results. But of course in practice the liability that banks create has monetary characteristics, and we therefore generate the well-defined demand in our models by way of a money demand function.

Two other important points need to be made concerning Tobin (1963). They relate to the fact that that paper's analysis, which is verbal rather than model-based, is essentially static and partial equilibrium, while the key arguments of the FMC view can only be understood using a dynamic and general equilibrium analysis. First, perhaps the most critical difference between FMC models, where banks can create their own funds, and ILF models, where they cannot, turns out to be that the variables on bank balance sheets, deposits and loans, are jump variables in the former case and predetermined variables in the latter. In economic terms, banks in the ILF model can only lend after attracting real savings, which can only be accumulated gradually over time, while banks in the FMC model can create new money instantaneously and independently of any available quantity of real aggregate saving. This difference affects exclusively the dynamics of the models, as the long-run steady states of both models are identical. But it has dramatic consequences for the economy's transition to its long-run steady state that simply cannot be captured in a static conceptual framework such as Tobin (1963). Second, Tobin (1963) relies in his arguments on the notion that when banks try to expand their balance sheets, some of their depositors will have more deposits than they wish to hold, and will switch to non-bank assets instead, thereby limiting the increase in the size of bank balance sheets. This may be true in some

[12] See for example Krugman (2013).

circumstances, but this in no way affects the size of the overall financial system. Because when deposits are withdrawn from the banking system, this can only happen if some matching assets are withdrawn from the same banking system. In the most extreme case of full withdrawal of all newly created deposits, the size of bank balance sheets does not change at all, but some assets that were previously held on bank balance sheets are now held on non-bank balance sheets. Overall, gross assets and liabilities throughout the economy have clearly increased by the exact amount of the bank loan. So long as the quantity of gross assets and liabilities is not neutral in the short and medium run, which is a key assumption of the FMC view with its emphasis on monetary effects, and which should have been an implicit assumption of Tobin (1963) with his emphasis on portfolio effects, this additional bank lending will therefore have real effects, irrespective of any partial equilibrium effects on bank balance sheets alone.

The analysis of Tobin (1963), and of the long subsequent literature in the same tradition[13], is therefore subject to the same critique that Culbertson (1958), Smith (1959) and others directed at Gurley and Shaw (1955, 1956). However, this debate did not continue much beyond the 1960s, as the macroeconomic and monetary functions of banks disappeared almost entirely from mainstream macroeconomic theory. As a result, many important insights of the past have been forgotten[14], and need to be relearned today.

2. ILF Models: Deposits Come Before Loans

The most basic, and also the most naïve, objection to a critique of the ILF view is that, surely, when I make a cheque deposit in a bank, the bank will use that deposit to fund loans to other households or firms. In other words, the bank intermediates my savings. What else would it do with "my money"? This objection exhibits both a confusion of microeconomic with macroeconomic arguments, and a confusion about the principles of double-entry bookkeeping. Figure 1 illustrates this with an example. In the four steps shown in that figure, a cheque with a value of 4 is deposited in Bank A, whose balance sheet is shown in the left column. But the deposited cheque, if it has any value, must be drawn on a deposit that already exists elsewhere in the banking system. In our example, it is drawn on Bank B, whose balance sheet is shown in the middle column. The right column shows the consolidated banking system, which is for simplicity assumed to consist of just Banks A and B. Also for simplicity, banks are assumed to have no net worth, and to keep central bank reserves of 10% against their deposits, much more than they would keep in practice.

The confusion of microeconomic and macroeconomic arguments becomes immediately obvious by considering the balance sheet of the consolidated banking system rather than of Bank A. It is entirely unaffected by this transaction. Deposits have been moved within the banking system, but this does not mean that the banking system as a whole has any more aggregate deposits to "fund loans". In a macroeconomic sense, this is clearly not what must be meant by the intermediation of savings.

But the fallacies go deeper than that. To begin, even Bank A does not have any additional funds to lend after it has received the deposit. At the moment the cheque is deposited, Bank A creates a new entry, the deposit, on the liabilities side of its balance sheet. But, by double-entry

[13] The reader is referred to Werner (2014a) for a comprehensive list of citations.
[14] One important exception is Werner (2005).

bookkeeping, there has to be a simultaneous matching entry elsewhere, which in this case is an accounts receivable entry on the asset side. This entry represents the liability of Bank B to deliver central bank reserves corresponding to the value of the cheque (this step is not shown in Figure 1). In other words, the funds are lent as soon as they are received – to Bank B. Bank A therefore has no additional funds to lend following the deposit.[15] The next step in Figure 1 is that Bank A sends the cheque for clearing, and clearing is settled using central bank reserves, with Bank B's central bank reserves decreasing by 4 and Bank A's reserves correspondingly increasing. One could now try to argue that Bank A can lend these additional central bank reserves to non-banks. But this would be a very elementary mistake. Central bank reserves simply cannot be lent to non-banks under the present split-circulation system, they are exclusively used to make payments between banks.[16] However, it might be argued, Bank A now has more reserves than it needs to support its deposit base, so there will be more lending by Bank A, and thus also more lending in aggregate. Notice that now we are no longer discussing lending by the bank of the funds represented by the original cheque deposit, because this is impossible, we are rather discussing indirect effects. But even this is incorrect. First, even if it was true that the additional reserves in Bank A cause it to lend more, Bank B faces the opposite situation, so it would lend less. We care about the aggregate outcome, which is unlikely to change because the overall quantity of reserves has not changed. Second, if Bank A cannot lend central bank reserves, and if it cannot create deposits through lending (under the ILF view of banking), how exactly can it lend more? Certainly not by attracting yet more deposits from Bank B, which will end up as yet more central bank reserves for Bank A, which cannot be lent. Bank A therefore, if it cannot create deposits through lending, has no ability to increase lending to non-banks after it receives the cheque deposit and the corresponding central bank reserves.

In the real world only fairly small settlement transactions in central bank reserves are typically required, because incoming and outgoing cheques approximately balance for Banks A and B. We nevertheless continue with our example. Given that Bank A does not need the additional central bank reserves to support its deposits with central bank liquidity, and because it cannot lend central bank reserves to non-banks, what it will do in the normal course of business is to lend them back to Bank A by way of an interbank loan. This is illustrated in the third row of Figure 1. Interbank loans are a way of reallocating central bank reserves to where they are most needed within the banking system. Once this transaction is complete, Bank A has therefore used the central bank reserves that came along with the additional deposit to make an interbank loan to Bank B. The deposit never enabled (or encouraged) it to lend more to non-banks, its only options were a loan of central bank reserves to Bank B or higher holdings of central bank reserves, which cannot be lent to non-banks.

A claim that a cheque deposit represents or leads to the intermediation of loanable funds is therefore a fallacy based on microeconomic or partial equilibrium arguments. But a large number of macroeconomic models exist in which banks do intermediate loanable funds in a general equilibrium setting. What do they have in mind? This is illustrated in Figure 2, which shows the story implicitly told by such models. Here we only need a single bank that represents the aggregate banking system. The story starts with the saver making a deposit. But we have just seen that this cannot be a cheque deposit.

[15] This is an excellent example of the critical importance of double-entry bookkeeping in the analysis of banking and finance - it keeps track of the full structure of gross claims and counterclaims that arise from financial transactions.

[16] The reason is that central bank reserves can only be held in accounts at the central bank, and the only institutions that can obtain such accounts are commercial banks and the central government.

It can also not be a cash deposit, for two reasons. First, cash is never "lent out", in the sense of a pure exchange of assets, loan against cash, on the bank's balance sheet. Cash can only be withdrawn against a pre-existing electronic deposit that has first been created in some other way. That other way is the subject of our inquiry here. Second, cash represents an extremely small fraction of the overall stock of money in modern economies, and banking transactions would proceed in exactly the way they proceed today if cash no longer existed at all. A model that would not be valid if this minor and non-constitutive element of our monetary system did not exist could therefore not be more than a theoretical exercise with no practical value.

It turns out that the only possible way to tell the story of ILF banks is that the saver makes a deposit of neither cheques nor cash but of *goods*. These goods must in turn have been accumulated through some combination of additional production of goods and foregone consumption of goods. A quick examination of the budget constraints used in modern general equilibrium models of banking shows that this is indeed, and to our knowledge almost without exception, the implicit assumption.

It is very important to try to understand what this would mean in practice, and we do so in Figure 2 by way of a concrete example. In this figure an agent called Saver approaches the bank to deposit a specific good that he happens to own, in this example gravel. In return the bank records a new deposit for Saver. At the moment of recording this deposit, by double-entry bookkeeping, the bank needs to record a matching entry elsewhere. This entry, on the asset side of its balance sheet, is an addition to its inventory of gravel. We now assume that an agent called Investor A[17] has approached the bank for a loan for the purpose of buying a machine, and that the bank has considered his proposal and decided to approve the loan. Continuing with our example, this loan must take the form of the bank exchanging the gravel against a loan contract with Investor A, in other words the loan is a portfolio swap on the asset side of the bank's balance sheet. Investor A drives away with gravel, and then negotiates a barter transaction with Investor B, whereby Investor B accepts the gravel in exchange for the new machine whose purchase Investor A wanted to finance. The bank is left with a deposit by Saver, and a loan to Investor A. It has intermediated loanable funds, in this example in the concrete form of gravel. These funds were the prerequisite for bank lending, and therefore for the physical investment of Investor A.

This story is fundamentally non-monetary, as the original bank deposit represents a receipt for goods, the loan represents a claim by the bank for future delivery of goods, and the ultimate purpose of the loan transaction can only be satisfied through barter of goods against goods. We are therefore left with a model where banks, who provide close to 100% of any modern economy's monetary medium of exchange, are modeled as institutions of barter.[18] Model economies that are constructed in this way are therefore entirely fictitious representations of reality, as such institutions simply do not exist.[19]

[17] To avoid misunderstandings, this agent is an entrepreneur, or an investor in real physical capital, rather than a financial asset investor.

[18] This is a point also emphasized by Graziani (1989, p.3): "... an economy using as money a commodity coming out of a regular process of production, cannot be distinguished from a barter economy."

[19] It could be argued that ILF models can nevertheless be used, because they can (perhaps) be calibrated to explain data as well as FMC models, even though the banks in such models have no counterpart in the real world. But this is logically comparable to insisting on the use of a Ptolemaic model of the solar system, based on the fact that Ptolemy's system managed to accurately track and predict some actual observations. Clearly no serious scientist would advocate this.

3. FMC Models: Loans Come Before Deposits

The FMC view of banks is illustrated in Figure 3. As in Figure 2, we only need a single bank that represents the aggregate banking system. This story does not start, but ends, with a saver making a deposit. It starts with a borrower, Investor A, approaching the bank for a loan - in the form of money, not goods. If the bank considers the credit risk of Investor A acceptable, it will enter into a loan contract. When the loan is entered into the bank's books as a new asset, a matching deposit is simultaneously entered as a new liability. The bank has created new purchasing power, money, through lending. Both the loan and the deposit are in the name of Investor A, which means that this transaction involves no intermediation of loanable funds whatsoever. Investor A now uses this new deposit to acquire a newly produced machine from Investor B, by transferring the new money in his account to the account of Investor B, in exchange for the machine. We assume for simplicity that Investor B leaves these funds as a deposit in the banking system. At this moment, Investor B becomes a saver. But what we want to emphasise is that Investor B's saving is a result, not a proximate cause, of the loan, and of the investment. As indicated in the passage from Schumpeter above, Investor B goes about his transaction with Investor A without any ex-ante intention of becoming a saver. His only intention is to sell machines, and to accept payment for his machines. In a modern economy cheques or money orders drawn on bank accounts are not only acceptable legal tender, they are the dominant practical means of making such payments, and Investor B would not remain in business for long if he did not accept them. But that means that he, or someone else to whom he might pass his deposit to make some business payments, has to end up being a new saver.

In many modern banking systems, loans to finance investment in the real economy have become a fairly small part of overall bank lending, with another part financing consumption, and a third and much larger part financing the exchange of existing real or financial assets between different agents (Hudson (2012)). If Investor B sold a pre-existing machine to Investor A, then his new deposit does not represent aggregate saving at all, rather it represents a portfolio exchange of his existing real asset against a new bank deposit. The absence of saving does not however make the bank loan any less essential, as the reallocation of assets only becomes possible because the bank creates new purchasing power for the use of the purchaser of the real asset.

The final balance sheet of the banking system is shown at the bottom of Figure 3. We find that, ex-post, the identity of the borrower, Investor A, is different from that of the depositor, Investor B. But this is not because the bank has intermediated real loanable funds from B to A, it is because it has created new purchasing power, exclusively for A, that was later transferred to B through the clearing system. As shown in the remainder of this paper, the mechanism through which this final balance sheet position is created is critically important, because the FMC and ILF mechanisms have very different macroeconomic implications.

Werner (2014a) shows empirically that the story told by Figure 3 is precisely what happens when a bank makes a new loan. He does so by tracing the entries created during the granting and disbursement of a new loan through a small bank's financial accounts. In addition, Werner (2014b) shows, in the UK context, that what distinguishes banks from non-banks, and therefore allows them to do this, is that they are exempt from legal rules known as Client Money Rules. These rules require non-banks to hold retail client monies in trust, or off-balance sheet, while banks are allowed to keep retail customer deposits on their own balance sheet. Depositors who deposit their money with a bank are therefore no longer the legal owners of this money, with the bank holding it in trust for them, but rather they are one of the general creditors of the bank. This implies that when non-banks disburse a loan to their clients, they need to give up either cash or their own bank deposits, while when banks disburse a loan, they do so by reclassifying an "accounts payable" liability (their obligation to disburse the loan in return for having received the right to receive future payments of principal and interest) as a "customer deposit".

B. DM Models? New Deposits Lead to Reserve Creation, Not Vice Versa

The DM view was widely accepted in academic and policymaking circles between the 1930s and the late 1960s[20], and therefore overlapped with the periods during which the FMC and ILF views dominated. In this section we cite leading policymakers and academics who have refuted the DM view, based on a combination of theoretical, institutional and empirical arguments.

The fact that the creation of broad monetary aggregates by banks comes prior to and in fact may (if commercial banks need more reserves) cause the creation of narrow monetary aggregates by the central bank is acknowledged in many descriptions of the money creation process by central banks and other policymaking authorities. The oldest and clearest comes from Alan Holmes (1969), who at the time was vice president of the New York Federal Reserve: "In the real world, banks extend credit, creating deposits in the process, and look for the reserves later." This is *exactly* the view put forward in this paper. Ulrich Bindseil (2004), at the time head of liquidity management at the European Central Bank: "It appears that with RPD [reserve position doctrine, i.e. the money multiplier theory] academic economists developed theories detached from reality, without resenting or even admitting this detachment." Charles Goodhart (2007), the UK's preeminent monetary economist: "... as long as the Central Bank sets interest rates, as is the generality, the money stock is a dependent, endogenous variable. This is exactly what the heterodox, Post-Keynesians ... have been correctly claiming for decades, and I have been in their party on this." Borio and Disyatat (2009), in a Bank for International Settlements working paper: "In fact, the level of reserves hardly figures in banks' lending decisions. The amount of credit outstanding is determined by banks' willingness to supply loans, based on perceived risk-return trade-offs and by the demand for those loans." Disyatat (2010), again from the BIS: "This paper contends that the emphasis on policy-induced changes in deposits is misplaced. If anything, the process actually works in reverse, with loans driving deposits. In particular, it is argued that the concept of the money multiplier is flawed and uninformative in terms of analyzing the dynamics of bank lending." Carpenter and Demiralp (2010), in a Federal Reserve Board working paper: "While the institutional facts alone provide compelling support for our view, we also demonstrate empirically that the relationships implied by the money multiplier do not exist in the data ... Changes in reserves are unrelated to changes in lending, and open market operations do not have a direct impact on lending. We conclude that the textbook treatment of money in the transmission mechanism can be rejected...". William C. Dudley (2009), president of the New York Federal Reserve Bank: "... the Federal Reserve has committed itself to supply sufficient reserves to keep the fed funds rate at its target. If banks want to expand credit and that drives up the demand for reserves, the Fed automatically meets that demand in its conduct of monetary policy." European Central Bank (2012), May 2012 Monthly Bulletin (emphasis added): "The occurrence of significant excess central bank liquidity does not, in itself, necessarily imply an accelerated expansion of ... credit to the private sector. If credit institutions were constrained in

[20] The reader is again referred to Werner (2014a) for a comprehensive list of citations.

their capacity to lend by their holdings of central bank reserves, then the easing of this constraint would result mechanically in an increase in the supply of credit. *The Eurosystem, however, ... always provides the banking system with the liquidity required to meet the aggregate reserve requirement.* In fact, the ECB's reserve requirements are backward-looking, i.e. they depend on the stock of deposits (and other liabilities of credit institutions) subject to reserve requirements as it stood in the previous period, and thus after banks have extended the credit demanded by their customers." Finally, academic critiques of the deposit multiplier model also exist (Kydland and Prescott (1990), Brunner and Meltzer (1990), Lombra (1992)), although recently this issue has received much less attention due to the disappearance of monetary aggregates from modern monetary models.

C. The Role of Policy

We conclude that a realistic macroeconomic model of the financial system has to reflect two facts. First, banks provide financing, meaning the creation of purchasing power through the creation of offsetting gross financial positions on their balance sheets. They do not intermediate real loanable funds, or savings. Second, banks' main constraint on the quantity of financing comes from parameters that enter their profit maximization problem, including most importantly shocks to their expectations of economic fundamentals. The availability of central bank reserves is not among these parameters. But the policy rate and regulatory requirements are.

In order for the policy rate, which affects the price of credit via arbitrage with other interest rates, to have a significant effect on the quantity of credit and money, it has to reach a point where the creditworthiness of borrowers is materially affected. McLeay et al. (2014a,b) argue that the effects of the policy rate on credit tend to go in this desired direction. But because the policy rate is generally assigned to controlling inflation, control of credit and money growth through this instrument tends to be weak and incidental. Altunbas, Gambacorta and Marques-Ibanez (2009) provide empirical evidence that confirms this for Europe. On the other hand, regulatory capital or liquidity requirements can potentially have very strong effects on credit growth, by affecting banks' incentives to lend in a much more targeted fashion than the policy rate.

III. Related Theoretical Literature

In relation to our paper, the recent literature[21] on financial frictions in macroeconomics can be divided into three groups. In the first group, all lending is direct and banks are absent. In the second and third groups, banks are present, but they are almost invariably modeled according to the ILF view.[22] In the second group, banks' net worth and balance sheets play no material role in the analysis (typically because all lending risk is diversifiable), and the emphasis is on loan pricing. In the third group, banks' balance sheets and net worth do play a role.

[21] An older literature on the credit channel view of monetary policy is summarised in Kashyap and Stein (1993) and Kashyap, Stein and Wilcox (1993). Partial equilibrium corporate finance models of banking will not be discussed in this paper.

[22] Apart from this paper and Benes and Kumhof (2012), we are aware of only two papers, Goodfriend and McCallum (2007) and Chari and Phelan (2014), that embrace the FMC view of banking.

We can be brief concerning the first group of models, where all lending is direct, because this is far removed from the topic of this paper. It includes one of the workhorse models of the modern literature, Kiyotaki and Moore (1997), where patient lenders extend direct credit against the real collateral offered by impatient borrowers. Iacoviello (2005) and Jermann and Quadrini (2012) also belong to this group.

The second group of models, where banks exist, but their balance sheets and net worth play no material role, is by far the largest. It includes another of the workhorse models of the modern literature, Bernanke, Gertler and Gilchrist (1999). In this model, which is also used in Christiano, Motto and Rostagno (2014), risk-neutral bankers make zero profits on their loans at all times. Capital adequacy regulation is therefore redundant, and bank net worth is absent from these models. The main function of banks is in generating an external financing premium for borrowers, which means that the focus of the analysis is on the balance sheet and leverage of the borrower, not of the bank. The same is true in Cúrdia and Woodford (2010), de Fiore, Teles and Tristani (2011) and Boissay, Collard and Smets (2013).

In the third group of models, which includes Gerali, Neri, Sessa and Signoretti (2010), Gertler and Karadi (2011), Gertler and Kiyotaki (2011), Goodhart, Kashyap, Tsomocos and Vardoulakis (2012), Adrian and Boyarchenko (2013) and Clerc, Derviz, Mendicino, Moyen, Nikolov, Stracca, Suarez and Vardoulakis (2014), bank balance sheets and net worth do matter, either through an incentive constraint under moral hazard or through a regulatory constraint. But banks are modeled according to the ILF view of banking.

We conclude this section by providing a complete list of the critical ingredients of FMC models of banking: *First*, banks do not intermediate pre-existing loanable funds in the form of goods, but create new deposits, in the form of money, through lending. *Second*, household demand for bank liabilities is modeled as a technology whereby bank deposits save transactions costs, rather than as a liquidity-free investment vehicle for real savings. *Third*, banks have their own balance sheet and net worth, and their net worth is endogenously determined through the interaction of non-diversifiable aggregate risk, financial market imperfections, macroprudential regulation and debt contracts. *Fourth*, banks are lenders and are therefore exposed to credit risk, but they are not holders of risky corporate equity and therefore exposed to price risk. *Fifth*, acquiring fresh bank capital from the equity markets is subject to market imperfections, as in Gertler and Karadi (2011). *Sixth,* bank lending is based on the loan contract of Bernanke et al. (1999), but with the difference that lending is risky due to non-contingent lending interest rates. *Seventh*, bank capital is subject to regulation that replicates features of the Basel regulatory framework, including minimum capital adequacy ratios whose violation results in penalties, and endogenously determined (through the interaction of all agents' optimisation problems) capital conservation buffers.

IV. The Models

A. Overview

To study the differences between ILF and FMC models of banking, we develop and simulate two pairs of models that each consist of one ILF model and one FMC model. Both model pairs illustrate important and distinct aspects of the difference between these model classes. In the first pair of models banks interact only with households, while in the second pair they interact with households and entrepreneurs. Importantly, the set of agents with whom banks interact represents the only difference between the four models, and the real steady states of all four models are exactly identical.

The common elements in all four models are a manufacturing/union sector that operates the economy's aggregate production technology, and that sets prices and wages subject to nominal rigidities, a government that finances government spending by levying lump-sum taxes, and that sets nominal interest rates according to a conventional inflation-forecast-based rule, and a banking sector that has retail deposit, wholesale, and retail lending functions.

The role of banks in the four models is shown schematically in Figure 4. In ILF Model 1 the banking sector intermediates real loanable funds, or goods, between a saver household and a borrower household, while in FMC Model 1 the banking sector creates new money for a single representative household. In ILF Model 2 the banking sector intermediates real loanable funds between a representative household and an entrepreneur, while in FMC Model 2 the banking sector creates new money for an entrepreneur, who then uses this money to acquire additional capital from the representative household. This pair of models has the advantage of being more directly comparable to many recent models in the literature.

We model the demand for bank deposits by way of a transactions cost technology, as in Schmitt-Grohe and Uribe (2004). This is critical only for the FMC models, but it is also done in the ILF models, in order to maintain the symmetry of steady states.

To simplify the analytical derivations, the banking sector in our model is divided into three subsectors, a retail deposit bank that issues the economy's medium of exchange, a retail lending bank that determines the terms of the loan contract, and a wholesale bank that ensures compliance with macroprudential regulations. Nominal and real interest rates on government debt are denoted by i_t and r_t, where $r_t = i_{t-1}/\pi_t^p$, where $\pi_t^p = P_t/P_{t-1}$, and where P_t is the GDP deflator. Wholesale lending rates are $i_{\ell,t}$ and $r_{\ell,t}$, retail lending rates, which add a credit risk spread to wholesale rates, are $i_{r,t}$ and $r_{r,t}$, and deposit rates are $i_{d,t}$ and $r_{d,t}$.

The model economy is assumed to be closed, and experiences constant positive technology growth $x = T_t/T_{t-1}$, where T_t is the level of labour augmenting technology. When the model's nominal variables, say X_t, are expressed in real normalised terms, we divide by P_t and by the level of technology T_t. We use the notation $\check{x}_t = X_t/\left(T_t P_t\right) = x_t/T_t$, with the steady state of $\check{x}_t$ denoted by $\bar{x}$.

We begin our exposition with the sectors that are common to all models, and end with a description of the special features of each of the four models. The exposition is kept brief in the interest of space.

B. Banking Sector

1. Retail Deposit Banks

Retail deposit banks have unit mass and are indexed by j, where individual banks differ by the deposit variety they offer. Retail deposit banks create deposit money $d_t(j)$ to purchase wholesale loans $o_t(j)$ and government bonds $b_t(j)$. Because wholesale loans are perfect substitutes for government bonds in the creation of deposit money, their interest rate is arbitraged with the policy rate, so that the nominal cost of lending by retail deposit banks to wholesale banks equals the policy rate. Retail depositors require a CES composite d_t of different deposit varieties, with elasticity of substitution σ, so that retail deposit banks act as monopolistic competitors vis-à-vis depositors. Letting $s = \sigma/(\sigma+1)$, we have an optimal price setting condition

$$i_{d,t} = s \cdot i_t \ . \tag{1}$$

Aggregate normalised profits of retail deposit banks are given by $\check{\Pi}_t^R = \left(\frac{r_t}{x} - \frac{r_{d,t}}{x}\right) \check{d}_{t-1}$. In equilibrium government debt will be zero at all times, so that wholesale loans are equal to retail deposits. We will therefore from now on, to simplify the exposition, set $\check{o}_t = \check{d}_t$, where we have dropped the index j because in equilibrium all retail deposit banks behave identically.

2. Wholesale Banks

Wholesale banks have unit mass and are indexed by j, where individual banks differ by the size of their balance sheet. Wholesale banks' nominal and real normalised loan stock between periods t and $t+1$ is given by $L_t(j)$ and $\check{\ell}_t(j)$, while their deposit stock is $D_t(j)$ and $\check{d}_t(j)$, and net worth is $N_t^b(j)$ and $\check{n}_t^b(j)$. Their balance sheet, in real normalised terms, is therefore given by

$$\check{\ell}_t(j) = \check{d}_t(j) + \check{n}_t^b(j) \ . \tag{2}$$

Because central bank reserves do not constrain the ability of banks to extend loans, banks are modeled as having no incentive, either regulatory or precautionary, to maintain cash reserves at the central bank. Because, furthermore, for households cash is dominated in return by bank deposits, in this economy there is no demand for government-provided real cash balances.

Banks are assumed to face pecuniary costs of falling short of official minimum capital adequacy ratios. The regulatory framework that we assume introduces a discontinuity in outcomes for banks. In any given period, a bank either remains sufficiently well capitalised, or it falls short of capital requirements and must pay a penalty. In the latter case, bank net worth suddenly drops further. The cost of such an event, weighted by the appropriate probability, is incorporated into the bank's optimal capital choice. Modeling this regulatory framework under the assumption of homogenous banks would lead to unrealistic outcomes where all banks simultaneously either pay or do not pay the penalty. We therefore instead assume a continuum of banks, each of which is exposed to idiosyncratic shocks. This implies that there is a continuum of ex-post capital adequacy ratios across banks, and a time-varying small fraction of banks that have to pay penalties in each period. This idiosyncratic risk can reflect a number of individual bank characteristics, such as differing success at raising non-interest income and minimising non-interest expenses, where the sum of the two equals zero over all banks.

Specifically, we assume that at the beginning of period $t+1$ each individual wholesale bank draws a shock ω^b_{t+1} such that the idiosyncratic return on its loan book equals $r_{\ell,t+1}\omega^b_{t+1}$, where ω^b_{t+1} is a unit mean lognormal random variable distributed independently over time and across banks. The standard deviation of $\ln(\omega^b_{t+1})$ equals $\left(\sigma^b_{t+1}\right)^2$, and the density function and cumulative density function of ω^b_{t+1} are denoted by $f^b(\omega^b_{t+1})$ and $F^b(\omega^b_{t+1})$.

The regulatory framework stipulates that banks have to pay a real penalty of $\chi\check{\ell}_t(j)$ at time $t+1$ if the sum of the gross returns on their loan book, net of gross deposit interest expenses and loan losses, is less than a fraction γ of the gross return on their loan book. In other words, a penalty is payable if at time $t+1$ net worth should be less than γ times the value of assets, so that γ can be interpreted as the Basel minimum capital adequacy ratio (MCAR).[23] Then the penalty cutoff condition is given by

$$r_{\ell,t+1}\check{\ell}_t(j)\omega^b_{t+1} - r_{t+1}\check{d}_t(j) + \check{\Pi}^R_{t+1}(j)x - \check{\Lambda}^\ell_{t+1}(j)x < \gamma r_{\ell,t+1}\check{\ell}_t(j)\omega^b_{t+1} \ . \tag{3}$$

The term $\check{\Pi}^R_{t+1}(j)$ represents the pro-rated lump-sum share received by bank j of the profits of retail deposit banks, and the term $\check{\Lambda}^\ell_{t+1}(j)$ represents the pro-rated lump-sum share paid (or received) by bank j of the net losses of retail lending banks, where the shares are pro-rated in proportion to each bank's net worth. We denote the cutoff idiosyncratic shock to loan returns below which the MCAR is breached ex-post by $\bar{\omega}^b_t$. Exploiting the fact that in equilibrium the ratios to net worth of loans, deposits, retail deposit profits and retail lending net losses are identical across all banks, we can write

$$\bar{\omega}^b_t \equiv \frac{r_t\check{d}_{t-1} - \check{\Pi}^R_t x + \check{\Lambda}^\ell_t x}{(1-\gamma)\, r_{\ell,t}\check{\ell}_{t-1}} \ . \tag{4}$$

Banks choose the volume of loans to maximise their pre-dividend net worth, which equals the gross return on loans, minus the gross cost of deposits, plus profits on retail deposit operations, minus net losses on retail lending operations, minus penalties on those banks that fall below the regulatory minimum:[24]

$$\underset{\check{\ell}_t(j)}{Max}\ E_t\left[r_{\ell,t+1}\check{\ell}_t(j)\omega^b_{t+1} - r_{t+1}\check{d}_t(j) + \check{\Pi}^R_{t+1}(j)x - \check{\Lambda}^\ell_{t+1}(j)x - \chi\check{\ell}_t(j)F^b\left(\bar{\omega}^b_{t+1}\right)\right] \ .$$

The optimality condition is

$$E_t\left[r_{\ell,t+1} - r_{t+1} - \chi\left(F^b\left(\bar{\omega}^b_{t+1}\right) + f^b\left(\bar{\omega}^b_{t+1}\right)\left(\frac{r_{t+1} + \frac{\check{\Pi}^R_{t+1}x}{\check{n}^b_t}}{(1-\gamma)\, r_{\ell,t+1}\frac{\check{\ell}_t}{\check{n}^b_t}}\right)\right)\right] = 0 \ . \tag{5}$$

This states that banks' wholesale lending rate is at a premium over the policy rate, by a margin that depends on the size of the MCAR γ, the penalty coefficient χ for breaching the MCAR, and expressions $F^b\left(\bar{\omega}^b_{t+1}\right)$ and $f^b\left(\bar{\omega}^b_{t+1}\right)$ that reflect the expected riskiness of banks σ^b_{t+1} and therefore the likelihood of a breach. Banks' retail lending rate, whose determination is discussed

[23] Note that in the model all assets have a risk-weighting of 100%, so that there is no difference between the Basel III capital adequacy ratio (which is calculated on the basis of risk-weighted assets) and the inverse of the Basel III leverage ratio (which is calculated on the basis of unweighted assets).

[24] As in Bernanke et al. (1999) and Christiano et al. (2014), our setup for banks, and also for their borrowers (see the next subsection), abstracts from the fact that their ultimate owners, households, have a variable intertemporal marginal rate of substitution whereby future profits are more valuable in some states of nature than in others.

in the next subsection, is at another premium over the wholesale lending rate, to compensate for the bankruptcy risk of borrowers. A sensible interpretation of the wholesale rate is therefore as the rate that a bank would charge to a hypothetical borrower (not present in the model) with zero default risk.

Another endogenous outcome of this optimisation problem is banks' actually maintained capital adequacy ratio γ_t^a. This will be considerably above the minimum requirement γ, because by maintaining an optimally chosen buffer banks protect themselves against the risk of penalties while minimising the cost of excess capital. There is no simple formula for γ_t^a, which in general depends nonlinearly on a number of parameters.

Banks' aggregate net worth $\check{n}_t^b$ is given by

$$\check{n}_t^b = \frac{1}{x}\left(r_{\ell,t}\check{\ell}_{t-1} - r_t\check{d}_{t-1} + \check{\Pi}_t^R x - \check{\Lambda}_t^\ell x - \chi\check{\ell}_{t-1}F_t^b\right) - \delta^b\check{n}_t^b \quad , \tag{6}$$

where $F_t^b = F^b\left(\bar{\omega}_t^b\right)$, and $\delta^b\check{n}_t^b$ are bank dividends, which are paid out to households in a lump-sum fashion. This specification of dividends, as explained in much more detail in Benes and Kumhof (2012), can be obtained by applying the "extended family" approach of Gertler and Karadi (2011).

3. Retail Lending Banks

Borrowers of retail lending banks have unit mass and are indexed by j, where individual borrowers differ by the size of their balance sheet. Each borrower uses an optimally chosen combination of bank loans $\check{\ell}_t(j)$ and internal funds to purchase physical capital $\check{k}_t(j)$ at the market price q_t. The financial return to capital is given by $ret_{k,t} = \left(q_t\left(1-\Delta\right) + r_{k,t}\right)/q_{t-1}$, where Δ is the physical depreciation rate and $r_{k,t}$ is the rental rate of capital. After the asset purchase, at the beginning of period $t+1$, each individual borrower draws a shock ω_{t+1}^k such that his idiosyncratic return to capital equals $ret_{k,t+1}\omega_{t+1}^k$, where ω_{t+1}^k is a unit mean lognormal random variable distributed independently over time and across borrowers. The standard deviation of $\ln(\omega_{t+1}^k)$, σ_{t+1}^k, is itself a stochastic process that will play a key role in our analysis. We will refer to it as the borrower riskiness shock. The density function and cumulative density function of ω_{t+1}^k are given by $f^k(\omega_{t+1}^k)$ and $F^k(\omega_{t+1}^k)$.

Each borrower receives a loan contract from the bank. This specifies a nominal loan amount $L_t(j)$, a gross nominal retail rate of interest $i_{r,t}$, payable as long as ω_{t+1}^k turns out to be sufficiently high to avoid default, and the fraction κ_t of the value of capital against which the bank is willing to lend. The most important difference between our model and those of Bernanke et al. (1999) and Christiano et al. (2014) is that the interest rate $i_{r,t}$ is assumed to be pre-committed in period t, rather than being determined in period $t+1$ after the realisation of time $t+1$ aggregate shocks.[25] The latter assumption ensures zero ex-post profits for banks at all times, while under our debt contract banks make zero expected profits, but realised ex-post profits generally differ from zero. Borrowers who draw ω_{t+1}^k below a cutoff level $\bar{\omega}_{t+1}^k$ cannot pay the interest rate $i_{r,t}$ and declare bankruptcy. They must hand over all their pledged assets, which exclude the fraction $(1-\kappa_t)$ against which banks did not lend, to the bank, but the bank can only recover a fraction

[25] See Bernanke et al. (1999): "... conditional on the ex-post realization of R_{t+1}^k, the borrower offers a (*state-contingent*) *non-default* payment that guarantees the lender a return equal in expected value to the riskless rate."

$(1-\xi)$ of the asset value of such borrowers. The remaining fraction represents monitoring costs. Retail lending banks' ex-ante zero profit condition, in real terms, is given by

$$E_t\left\{r_{\ell,t+1}\check{\ell}_t(j)-\left[\left(1-F^k(\bar{\omega}^k_{t+1})\right)r_{r,t+1}\check{\ell}_t(j)+(1-\xi)\int_0^{\bar{\omega}^k_{t+1}}\kappa_t q_t\check{k}_t(j)ret_{k,t+1}\omega^k f^k(\omega^k)d\omega^k\right]\right\}=0\ .$$

This states that the expected payoff to lending must equal wholesale interest charges $r_{\ell,t+1}\check{\ell}_t(j)$. The first term in square brackets is the gross real interest income on loans to borrowers whose idiosyncratic shock exceeds the cutoff level, $\omega^k_{t+1}\geq\bar{\omega}^k_{t+1}$. The second term is the amount collected by the bank in case of the borrower's bankruptcy, where $\omega^k_{t+1}<\bar{\omega}^k_{t+1}$. This cash flow is based on the return $ret_{k,t+1}\omega^k$ on the purchase value of capital $q_t\check{k}_t(j)$, but multiplied by two additional factors. First, the factor κ_t represents the fraction of the value of underlying capital against which the bank, at the time of setting its lending rate, is willing to lend, and which it is therefore able to recover in a bankruptcy. Second, the factor $(1-\xi)$ contains a proportional bankruptcy cost ξ that the bank loses when recovering the value of low-return projects.

The ex-post cutoff productivity level is determined by equating, at $\omega^k_t=\bar{\omega}^k_t$, the gross interest charges payable by the borrower in the event of continuing operations $r_{r,t}\check{\ell}_{t-1}(j)$, in other words the cost of not defaulting, to the gross idiosyncratic return on the borrower's asset that needs to be handed over to the bank in the event of not continuing operations, $ret_{k,t}\kappa_{t-1}q_{t-1}\check{k}_{t-1}(j)\bar{\omega}^k_t$, in other words the cost of defaulting. Exploiting the fact that in equilibrium the ratios to internal funds of assets and loans are identical across all borrowers, we can write

$$\bar{\omega}^k_t=\frac{r_{r,t}\check{\ell}_{t-1}}{ret_{k,t}\kappa_{t-1}q_{t-1}\check{k}_{t-1}}\ . \tag{7}$$

We denote, following Bernanke et al. (1999) and Christiano et al. (2014), the lender's gross share in pledged[26] assets' earnings by $\Gamma_{t+1}=\Gamma(\bar{\omega}^k_{t+1})$, and the lender's monitoring costs share in pledged assets' earnings by $\xi G_{t+1}=\xi G(\bar{\omega}^k_{t+1})$.[27] The borrower is left with a share $1-\kappa_t\Gamma_{t+1}$ of total assets' earnings. Then (7) can be used to express the zero profit condition of banks in a way that determines the retail lending rate:

$$E_t\left\{\left(1-F^k(\bar{\omega}^k_{t+1})\right)\frac{r_{r,t+1}}{r_{\ell,t+1}}+(1-\xi)\,G_{t+1}\frac{ret_{k,t+1}}{r_{\ell,t+1}}\frac{\kappa_t q_t\check{k}_t}{\check{\ell}_t}\right\}=1\ . \tag{8}$$

In other words, the bank will set the unconditional lending rate such that its expected earnings are sufficient to cover the opportunity cost of the loan plus monitoring costs.

The remainder of the analysis is similar to Bernanke et al. (1999), except for the fact that the lending rate is not conditional on period $t+1$ shock realizations, and for the presence of the willingness-to-lend coefficient κ_t. Specifically, the borrower selects the optimal level of investment by maximizing $E_t\left\{(1-\kappa_t\Gamma_{t+1})\,q_t\check{k}_t ret_{k,t+1}\right\}$, the expected net return on capital, subject to (8). The conditions for the optimal loan contract differ depending on whether the borrower is a household (ILF Model 1 and FMC Model 1) or an entrepreneur (ILF Model 2 and FMC Model 2), and will therefore be deferred until the problems of these agents are discussed.

[26] The term "pledged" refers to the fraction κ_t of assets against which banks lent.

[27] The full expressions are $\Gamma(\bar{\omega}^k_{t+1})=\int_0^{\bar{\omega}^k_{t+1}}\omega^k_{t+1}f^k(\omega^k_{t+1})d\omega^k_{t+1}+\bar{\omega}^k_{t+1}\int_{\bar{\omega}^k_{t+1}}^{\infty}f^k(\omega^k_{t+1})d\omega^k_{t+1}$ and $\xi G(\bar{\omega}^k_{t+1})=\xi\int_0^{\bar{\omega}^k_{t+1}}\omega^k_{t+1}f^k(\omega^k_{t+1})d\omega^k_{t+1}$.

Retail lending banks' net loan losses $\check{\Lambda}_t^\ell$ are positive if wholesale interest expenses, which are the opportunity cost of retail lending banks' lending, exceed banks' net (of monitoring costs) share in borrowers' gross earnings on pledged assets. This will be the case if a larger than anticipated number of borrowers defaults, so that, ex-post, banks find that they have set their pre-committed retail lending rate at an insufficient level to compensate for lending losses. Of course, if losses are positive for banks, this corresponds to gains for their borrowers. Banks' ex-post loan losses are given by

$$\check{\Lambda}_t^\ell x = r_{\ell,t}\check{\ell}_{t-1} - \kappa_{t-1}q_{t-1}\check{k}_{t-1}ret_{k,t}\left(\Gamma_t - \xi G_t\right) \ . \tag{9}$$

C. Manufacturing Sector

1. Manufacturers

Manufacturers have unit mass and are indexed by j, where individual manufacturers differ by the goods variety that they produce and sell. They purchase an aggregate of labour services $h_t(j)$ from unions, at the aggregate producer nominal wage rate V_t, and capital services $k_{t-1}(j)$ from households (in ILF Model 1 and FMC Model 1) or entrepreneurs (in ILF Model 2 and FMC Model 2), at the nominal rental rate R_t^k. Their differentiated output $y_t(j)$ is sold, at price $P_t(j)$, for the purpose of consumption, investment, government spending, monitoring activities and monetary transactions costs. In each case, demand is for a CES aggregate over individual output varieties, with elasticity of substitution θ_p, and thus with a gross mark-up that equals $\mu_p = \theta_p/(\theta_p - 1)$. The production function of an individual manufacturer is given by

$$y_t(j) = \left(T_t h_t(j)\right)^{1-\alpha} k_{t-1}(j)^\alpha \ . \tag{10}$$

Optimality conditions for cost minimization are standard. Each manufacturer faces price adjustment costs that are quadratic in changes in the rate of price inflation:

$$\check{G}_{P,t}(j) = \frac{\phi_p}{2}\check{y}_t\left(\frac{\pi_t^p(j)}{\pi_{t-1}^p} - 1\right)^2 \ . \tag{11}$$

This is similar to the sticky price inflation formulation first introduced by Ireland (2001). The optimality condition for price setting is a standard New Keynesian Phillips curve. For future reference, manufacturer profits are denoted by $\check{\Pi}_t^M$.

2. Unions

Unions have unit mass and are indexed by j, where individual unions differ by the labour variety that they sell. Specifically, unions purchase homogenous labour services from households, at the household nominal wage rate W_t, and sell differentiated labour varieties to manufacturers, at the union-specific producer wage $V_t(j)$. Manufacturers demand a CES aggregate over individual labour varieties, with elasticity of substitution θ_w, and thus with gross mark-up $\mu_w = \theta_w/(\theta_w - 1)$. Each union faces wage adjustment costs that are quadratic in changes in the rate of wage inflation:

$$\check{G}_{W,t}(j) = \frac{\phi_w}{2}h_t\left(\frac{\pi_t^w(j)}{\pi_{t-1}^w} - 1\right)^2 \ . \tag{12}$$

The optimality condition is a standard New Keynesian Phillips curve for wage setting. For future reference, union profits are denoted by $\check{\Pi}_t^U$.

D. ILF Model 1: Saver and Borrower Households

For this first model variant, the remainder of the economy consists of two household groups, savers (superscript s) with population share ϖ and borrowers (superscript b) with population share $1-\varpi$. Because we express all decision variables of savers and borrowers in per capita terms, all of the optimality conditions in Sections IV.B and IV.C, if they contain those decision variables, need to be amended to include the respective weights.[28] This is only true for ILF Model 1, all subsequent models contain only one representative household and, except for the special features of those models, the above optimality conditions are complete as stated.

Letting $j \in \{s,b\}$, both savers and borrowers consume, c_t^j, and supply labour, h_t^j. Aggregate consumption and labour supply are given by $c_t = \varpi c_t^s + (1-\varpi)\, c_t^b$ and $h_t = \varpi h_t^s + (1-\varpi)\, h_t^b$. Household preferences are

$$Max \quad E_0 \sum_{t=0}^{\infty} \beta^t \left\{ S_t^c (1-\frac{v}{x}) \log(c_t^j - v\bar{c}_{t-1}^j) - \psi^j \frac{h_t^j(i)^{1+\frac{1}{\eta}}}{1+\frac{1}{\eta}} \right\} , \tag{13}$$

where $\bar{c}_t^j$ is aggregate per capita consumption for household group j. Note that the only parameter that differs across household types is ψ^j, a scale parameter that is used to normalise steady state labour supplies. All preference parameters that affect model dynamics, β, v and η, are identical across savers and borrowers, and will retain the same values in model variants with a representative household. This helps to guarantee that the steady states of all four models are identical. The equality of discount factors β is worth stressing. Typical ILF models feature patient savers and impatient borrowers. However, in models where bank liabilities are held for their monetary services rather than as a saving instrument, there is no necessary correlation between the status of an agent as a bank depositor and greater patience.

1. Saver Household

Deposit money balances for consumption and investment purposes $\check{d}_t^c$ and $\check{d}_t^i$ are held exclusively by saver households.[29] We adopt the money demand specification of Schmitt-Grohe and Uribe (2004). Specifically, defining velocities as $v_t^c = \check{c}_t^s / \check{d}_t^c$ and $v_t^i = \check{I}_t / \check{d}_t^i$, and with $j \in \{c,i\}$, we have proportional transactions costs of[30]

$$s_t^j = A^j v_t^j + \frac{B^j}{v_t^j} - 2\sqrt{A^j B^j} \ . \tag{14}$$

At the beginning of each period, the representative saver household splits into two groups, consumers/workers and capital goods producers. We define a lump-sum net income stream Ω_t (in aggregate rather than per capita terms) that consists of firm dividends, plus payments related to

[28] Specifically, the borrower-specific variables are $\check{k}_t$ and $\check{\ell}_t$, so that $\check{k}_t$ becomes $(1-\varpi)\,\check{k}_t$, and $\check{\ell}_t$ becomes $(1-\varpi)\,\check{\ell}_t$. The saver-specific variables are $\check{I}_t$ and $\check{d}_t$, so that $\check{I}_t$ becomes $\varpi \check{I}_t$, and $\check{d}_t$ becomes $\varpi \check{d}_t$.

[29] It would be trivial to allow borrowers to hold some money balances. We chose not to do this in order to maintain the clean separation between one group of agents that lends to the banking system and another, separate group of agents that borrows from the banking system. This separation is characteristic of all ILF models of banking.

[30] While these money demand functions are commonly used in the literature, we apply them here to very broad monetary aggregates and to a novel model environment. Further work on the appropriate specification and calibration of money demand functions within this model class is a very important area for further research.

adjustment costs, minus government lump-sum taxes. We assume that savers and borrowers receive shares ι and $1-\iota$ of this income. Letting $\check{d}_t = \check{d}_t^c + \check{d}_t^i$, the budget constraint of saver households is given by

$$\begin{aligned} \check{d}_t &= \frac{r_{d,t}}{x}\check{d}_{t-1} + \check{w}_t h_t^s + \frac{\tilde{h}_t^s}{\tilde{h}_t}\left(\check{\Pi}_t^U + \check{G}_{W,t}\right) - \check{c}_t^s\left(1+s_t^c\right) \\ &\quad + \left(q_t - 1 - s_t^i\right)\check{I}_t - \check{G}_{I,t} + \frac{\iota}{\varpi}\check{\Omega}_t \ . \end{aligned} \tag{15}$$

Here $\tilde{h}_t^s$ and $\tilde{h}_t$ refer to aggregate hours of all saver households and of all households, which are taken as given by an individual household. The term $\check{G}_{I,t}$ represents investment adjustment costs, with functional form

$$\check{G}_{I,t} = \frac{\phi_i}{2}\overset{\vee}{\check{I}}_t\left(S_t^i\frac{\check{I}_t}{\check{I}_{t-1}} - 1\right)^2 \ , \tag{16}$$

where $\overset{\vee}{\check{I}}_t$ is aggregate per capita investment, again taken as given by the household. Finally, $\check{\Omega}_t$ is given by

$$\check{\Omega}_t = \delta^b \check{n}_t^b + \check{\Pi}_t^M + \check{G}_{P,t} + \varpi\check{G}_{I,t} - \check{\tau}_t \ , \tag{17}$$

where $\check{\tau}_t$ represents government lump-sum taxes. The first-order condition for labour supply is standard. The first-order conditions for money demands, consumption and investment contain monetary wedges whereby the intertemporal marginal rate of substitution is equated to $(1 - s_t^{j\prime}(v_t^j)^2)/r_{d,t+1}$, the effective price of consumption equals $1 + s_t^c + s_t^{c\prime}v_t$, and the effective price of investment equals $1 + s_t^i + s_t^{i\prime}v_t^i$. The effective prices of consumption and investment are decreasing in the amount of monetary purchasing power in circulation. The intuition is that more liquidity makes it less costly to purchase or sell consumption and investment goods.

2. Borrower Household

The economy's capital stock is exclusively held by the representative borrower household, who borrows against the collateral of this capital stock from the bank. The borrower's time t budget constraint is

$$\begin{aligned} & q_t\check{k}_t - \check{\ell}_t \\ =\ & q_{t-1}\check{k}_{t-1}\frac{ret_{k,t}}{x}\left(1 - \xi\kappa_{t-1}G_t\right) - \frac{r_{\ell,t}}{x}\check{\ell}_{t-1} + \check{\Lambda}_t^\ell \\ & + \check{w}_t h_t^b + \frac{\tilde{h}_t^b}{\tilde{h}_t}\left(\check{\Pi}_t^U + \check{G}_{W,t}\right) - \check{c}_t^b + \frac{1-\iota}{1-\varpi}\check{\Omega}_t \ , \end{aligned} \tag{18}$$

while for the expected time $t+1$ budget constraint the second line, relating to capital investment and borrowing, is replaced by $E_t\left(1 - \kappa_t\Gamma_{t+1}\right)\frac{ret_{k,t+1}}{x}q_t\check{k}_t$. The first-order condition for labour supply is standard. The first-order condition for consumption differs from that for saver households by the absence of the monetary wedge. The optimality conditions for capital and loans can be combined to yield a condition that is (except for the presence of κ_t) identical to that derived from more conventional household-entrepreneur setups, including Christiano et al. (2014) and also our own ILF Model 2 below:

$$E_t\left\{\frac{ret_{k,t+1}}{r_{\ell,t+1}}\left(1 - \kappa_t\Gamma_{t+1}\right) + \tilde{\lambda}_{t+1}\left(\frac{ret_{k,t+1}}{r_{\ell,t+1}}\kappa_t\left(\Gamma_{t+1} - \xi G_{t+1}\right) - 1\right)\right\} = 0 \ . \tag{19}$$

Here the variable $\tilde{\lambda}_{t+1}$ equals $\Gamma^{\omega}_{t+1}/\left(\Gamma^{\omega}_{t+1} - \xi G^{\omega}_{t+1}\right)$, where Γ^{ω}_{t+1} and G^{ω}_{t+1} are the partial derivatives of Γ_{t+1} and G_{t+1} with respect to $\bar{\omega}^{k}_{t+1}$, and where $\tilde{\lambda}_{t+1}$ represents an indicator of the tightness of bank lending conditions.

E. FMC Model 1: Representative Household

For this model variant, the remainder of the economy consists of a single representative household that both borrows from and holds deposits at the bank. Similar to saver households above, at the beginning of each period the representative household splits into two groups, consumers/workers/capital holders and capital goods producers.

The preferences of the representative household are identical to (13), after dropping all superscripts j. The household budget constraint is given by

$$\begin{aligned} & q_t\check{k}_t - \check{\ell}_t + \check{d}_t \qquad (20) \\ = \; & q_{t-1}\check{k}_{t-1}\frac{ret_{k,t}}{x}\left(1 - \xi\kappa_{t-1}G_t\right) - \frac{r_{\ell,t}}{x}\check{\ell}_{t-1} + \check{\Lambda}^{\ell}_t \\ & + \frac{r_{d,t}}{x}\check{d}_{t-1} + \left(q_t - 1 - s^i_t\right)\check{I}_t - \check{G}_{I,t} - \check{c}_t\left(1 + s^c_t\right) + \check{w}_t h_t + \check{\Omega}_t \;\; , \end{aligned}$$

while, as for the borrower household above, for the expected time $t+1$ budget constraint the second line is replaced by $\left(1 - \kappa_t\Gamma_{t+1}\right)\frac{ret_{k,t+1}}{x}q_t\check{k}_t$. First-order optimality conditions for consumption, investment and bank deposits are identical to those of the saver household in ILF Model 1, while those for capital and loans are identical to those of the borrower household in ILF Model 1.

The crucial difference between ILF Model 1 and FMC Model 1 is therefore found exclusively in the budget constraints of banks' customers, where the separate constraints (15) and (18) of ILF Model 1 become a single constraint (20) in FMC Model 1. This implies that, while deposits and loans are predetermined variables in ILF Model 1, representing slow-moving real savings, they are jump variables in FMC Model 1, representing fast-moving financing, created through matching gross positions on the balance sheets of banks.

It may at first sight be surprising that the difference between two radically different classes of banking models should be found exclusively in the optimisation problems of banks' customers, while the optimisation problems of banks in both models are formally identical. However, this is misleading. First, the critical feature of banks in the two model classes is the function they perform for their customers, namely intermediation of loanable funds between different customers (ILF) or financing through money creation for a single customer (FMC). This must be reflected in different optimisation problems for banks' customers. Second, the fact that banks' optimisation problems are formally identical is attributable to the fact that our models are designed to have identical steady states. This means that our ILC models feature deposits that enter transactions cost technologies, which is otherwise an unnecessary feature of ILC models. Removing it would make the optimisation problems of banks in the two model classes different.

F. ILF Model 2: Representative Household and Entrepreneur

With a single exception, ILF Model 2 is identical to FMC Model 1. The exception is that the representative household does not hold, and borrow against, the capital stock. Instead, a representative entrepreneur with a separate balance sheet performs these activities. The budget constraint (20) is therefore split into two separate budget constraints, one for the household, equation (21), and one for the entrepreneur, equation (22):

$$\check{d}_t = \frac{r_{d,t}}{x}\check{d}_{t-1} + \left(q_t - 1 - s_t^i\right)\check{I}_t - \check{G}_{I,t} - \check{c}_t\left(1 + s_t^c\right) + \check{w}_t h_t + \check{\Omega}_t \ , \tag{21}$$

$$q_t\check{k}_t - \check{\ell}_t = q_{t-1}\check{k}_{t-1}\frac{ret_{k,t}}{x}\left(1 - \xi\kappa_{t-1}G_t\right) - \frac{r_{\ell,t}}{x}\check{\ell}_{t-1} + \check{\Lambda}_t^{\ell} - \delta^k\left(q_t\check{k}_t - \check{\ell}_t\right) \ . \tag{22}$$

Here entrepreneur net worth equals $\check{n}_t^k = q_t\check{k}_t - \check{\ell}_t$, and $\delta^k\check{n}_t^k$ represents entrepreneur dividends. The combined optimality condition for capital and loans of the entrepreneur is exactly identical to the corresponding condition of the representative household. The key difference between this ILF model and the previous FMC model is therefore again not found in the optimality conditions but in the budget constraints. Specifically, in ILF Model 2, as in ILF Model 1, loans and deposits are predetermined variables.

G. FMC Model 2: Representative Household, Entrepreneur, Traded Capital

With a single exception, FMC Model 2 is identical to ILF Model 2. The exception is that the representative household and the representative entrepreneur can trade physical capital against bank deposits among themselves. Note that the bank's initial transaction remains with the borrower alone, in that its new loan to the entrepreneur is matched by a new deposit in the name of the same entrepreneur. This deposit can subsequently be transferred to the household, but this does not mean that the household creates the deposit through saving. A saving decision never enters this transaction, the household's only decision is to accept the deposit in payment for capital.

FMC Model 2 has a new market clearing condition for capital $\check{k}_t = \check{k}_t^e + \check{k}_t^h$, where $\check{k}_t^e$ is held by entrepreneurs and $\check{k}_t^h$ by households. The capital stock $\check{k}_t^h$, instead of entering into the production function of the manufacturer, derives a return from a separate technology with a lower steady state rate of return. In all optimality and equilibrium conditions of banks and manufacturers, the variable $\check{k}_t$ therefore needs to be replaced by $\check{k}_t^e$. The capital stock held by the representative household has a physical return $r_{h,t} = \bar{r}_h - \epsilon_h(\check{k}_{t-1}^h/x)$, where ϵ_h is a small elasticity parameter, and where we calibrate $\bar{r}_h$ so that in a steady state with $\bar{k}^h = 0$ the household is indifferent between holding physical capital and deposits. This ensures that the steady state of this model remains exactly identical to all previous models. The financial return to capital held by households is $ret_{h,t} = \left(q_t\left(1 - \Delta\right) + r_{h,t}\right)/q_{t-1}$. The time t budget constraints of household and entrepreneur (21) and (22) are therefore replaced by

$$\begin{aligned} & \check{d}_t + q_t\check{k}_t^h \\ = \ & \frac{r_{d,t}}{x}\check{d}_{t-1} + q_{t-1}\check{k}_{t-1}^h\frac{ret_{h,t}}{x} + \left(q_t - 1 - s_t^i\right)\check{I}_t - \check{G}_{I,t} - \check{c}_t\left(1 + s_t^c\right) + \check{w}_t h_t + \check{\Omega}_t \ , \end{aligned} \tag{23}$$

$$\begin{aligned} & q_t\check{k}_t^e - \check{\ell}_t \\ = \ & q_{t-1}\check{k}_{t-1}^e\frac{ret_{k,t}}{x}\left(1 - \xi\kappa_{t-1}G_t\right) - \frac{r_{\ell,t}}{x}\check{\ell}_{t-1} + \check{\Lambda}_t^{\ell} - \delta^k\left(q_t\check{k}_t^e - \check{\ell}_t\right) \ . \end{aligned} \tag{24}$$

The ability of households and entrepreneurs to trade physical capital against bank deposits implies that deposits and loans are jump variables, representing fast-moving financing rather than slow-moving saving, as in FMC Model 1. This is despite the fact that, unlike in FMC Model 1, two groups of non-bank agents, with separate net worth, interact with the bank.

H. Government and Market Clearing

Monetary policy follows a conventional inflation-forecast-based interest rate rule

$$i_t = (i_{t-1})^{m_i} \left(\frac{x\left(1 - \bar{s}^{c'} (\bar{v}^c)^2\right) \bar{\pi}}{\beta s} \right)^{(1-m_i)} \left(\frac{\pi_{4,t+3}}{(\bar{\pi})^4} \right)^{(1-m_i)m_\pi} , \tag{25}$$

where $\pi_{4,t} = \pi_t \pi_{t-1} \pi_{t-2} \pi_{t-3}$, and where the expression for the steady state nominal interest rate $(x(1 - \bar{s}^{c'} (\bar{v}^c)^2)\bar{\pi})/\beta s$ follows from the optimality conditions for deposits combined with the pricing condition of retail deposit banks. Government spending is assumed to equal a fixed fraction s_g of steady state output $\bar{y}$, $\check{g}_t = s_g \bar{y}$, and the government budget is balanced in each period by way of lump-sum taxes, $\check{\tau}_t = \check{g}_t$. Assuming that initial government debt equals zero, government debt therefore remains at zero at all times.

For ILF Model 1, ILF Model 2, and FMC Model 1, the goods market clearing condition is given by

$$\check{y}_t = \check{c}_t + \check{I}_t + \check{g}_t + \check{\mathcal{M}}_t^b + \check{\mathcal{M}}_t^k + \check{\mathfrak{T}}_t^c + \check{\mathfrak{T}}_t^i \ , \tag{26}$$

except that in ILF Model 1 $\check{I}_t$ is replaced by $\varpi \check{I}_t$. $\check{\mathcal{M}}_t^b$ are regulatory penalties on banks, $\check{\mathcal{M}}_t^k$ are costs of monitoring entrepreneurs, and $\check{\mathfrak{T}}_t^c + \check{\mathfrak{T}}_t^i$ are monetary transactions costs, which for model variants with representative households equal $\check{\mathfrak{T}}_t^c = \check{c}_t s_t^c$ and $\check{\mathfrak{T}}_t^i = \check{I}_t s_t^i$. For FMC Model 2 aggregate output, on the left-hand side of (26), has an additional component $r_{h,t} \left(\check{k}_{t-1}^h / x \right)$.

I. Shocks

We study the response of the four model economies to two financial shocks. We do this because financial shocks best illustrate the differences between ILF and FMC models, and also because such shocks are empirically important – both Jermann and Quadrini (2012) and Christiano et al. (2014) find that they can account for up to half of U.S. output volatility.[31] But we emphasise that the differences between ILF and FMC models remain very large for non-financial shocks.

The first shock, to σ_t^k, is the "risk shock" of Christiano et al. (2014). We refer to this as the borrower riskiness shock:

$$\log\left(\sigma_t^k\right) = (1 - \rho_z) \log\left(\bar{\sigma}^k\right) + \rho_z \log\left(\sigma_{t-1}^k\right) + \varepsilon_t^z \ . \tag{27}$$

The second shock, to κ_t, is to the fraction of the value of the capital stock against which banks are willing to lend. We refer to this as the willingness-to-lend shock:

$$\log(\kappa_t) = (1 - \rho_k) \log(\bar{\kappa}) + \rho_k \log(\kappa_{t-1}) + \varepsilon_t^k \ . \tag{28}$$

[31] This is in line with the writings of Fisher (1936) and many other leading economists of the 1930s and 1940s, who argued that fluctuations in banks' willingness to lend are a key driver of business cycle volatility.

The borrower riskiness shock is a shock to economic fundamentals, it is essentially a technology shock, and banks in the model rationally react to changes in these fundamentals. By contrast, with the willingness-to-lend shock the driving force is an autonomous change in banks' sentiment that is, in the model, not a function of other economic fundamentals. It is interesting to explore the behaviour of the model economy under such sentiment shocks, because it can often be hard to discern more fundamental reasons for economic booms and crashes (Carvalho, Martin and Ventura (2012), Martin and Ventura (2012)).

J. Calibration

The calibration of the model is based on U.S. historical data and on the DSGE literature insofar as it relates to the United States. For balance sheet and spreads data, we focus on the period prior to the Great Recession, thereby excluding the highly volatile data sample during the financial crisis. We emphasise that our calibration is illustrative rather than representing a full-blown econometric exercise. This is sufficient to clearly exhibit the fundamentally different behaviour of ILF and FMC models of banking.

We first describe our calibration for the simplest model variant, FMC Model 1, and then briefly comment on the other three models. The latter only contain a small number of additional dynamic parameters, while the steady state in each case remains identical to FMC Model 1.

We calibrate both the steady state real growth rate and the steady state inflation rate at 2% per annum, and the model's risk-free real interest rate at 3% per annum. The parameter α is calibrated to obtain a steady state labour share of 60%. This is in line with recent Bureau of Labor Statistics (BLS) data for the U.S. business sector (Ueda and Brooks (2011)). This share has exhibited a declining trend over recent decades, and we therefore base our calibration on the more recent values. The private investment to GDP ratio is set to 20% of GDP, roughly its average in U.S. data. The implied depreciation rate Δ, at around 8% per annum, is in the middle of the range of values typically used in the literature. The government spending to GDP ratio is set to its approximate historical average of 18% of GDP, by fixing the parameter s_g.

The calibration of household preferences is close to the related literature, with habit persistence at $v = 0.75$ and labour supply elasticity at $\eta = 1$. The parameter ψ is calibrated to normalise steady state labour supply to 1. The steady state price and wage mark-ups of monopolistically competitive manufacturers and unions are fixed, in line with much of the New Keynesian literature, at 10%, or $\mu_p = \mu_w = 1.1$. Wage and price stickiness parameters are calibrated as $\phi_w = 200$ and $\phi_p = 200$, which corresponds to average contract lengths of five quarters in a Calvo (1983) model with full indexation to past inflation. The monetary policy rule is calibrated at $m_i = 0.7$ and $m_\pi = 2.5$. The investment adjustment cost parameter is calibrated at $\phi_i = 0.5$ for ILF Model 1 and FMC Model 1, and at $\phi_i = 2.5$ for ILF Model 2 and FMC Model 2, as this generates similar investment responses in these two model groups. The choice of $\phi_i = 2.5$ in ILF Model 2 and FMC Model 2 follows Christiano, Eichenbaum and Evans (2005).

For the lending technology, we calibrate the parameter $\bar{\sigma}^k$ to yield a steady state spread of the retail lending rate $i_{r,t}$ over the policy rate i_t of 1.5%, consistent with the evidence in Ashcraft and Steindel (2008) and the discussion in Benes and Kumhof (2012). The parameter ξ is calibrated to produce a quarterly loan default rate of 1.5%, consistent with the evidence in Ueda and Brooks (2011). The parameters A^i and A^c determine the overall demand for bank liabilities and therefore the size of bank balance sheets. We use them to calibrate the steady state share of investment-related deposits in overall deposits at 50%, and the steady state leverage ratio, defined as loans divided by the difference between the value of physical capital and loans, at 100%. The latter is consistent with the evidence in Ueda and Brooks (2011).[32] The interest semi-elasticity of money demand is the percent change in money demand in response to a one percentage point increase in the opportunity cost of holding money. We set the money demand parameters B^j, $j \in \{c, i\}$, to obtain steady state interest semi-elasticities of money demand of 0.05, based on the estimates in Ball (2001).

For the banking sector, we calibrate bank riskiness σ^b such that the percentage of banks violating the minimum capital adequacy ratio equals 2.5% of all banks per quarter. The parameter determining the Basel minimum capital adequacy ratio is set to 8% of assets, $\gamma = 0.08$, as under both the Basel II and the new Basel III regulations. The parameter δ^b is calibrated to be consistent with the assumption that banks maintain an average actual capital adequacy ratio $\bar{\gamma}^a$ of 10.5%, which means that they maintain a capital conservation buffer of 2.5%, again as envisaged under Basel III.

Together with our assumptions about household leverage and money demand, this implies an overall volume of bank lending for physical investment purposes equal to 120% of GDP. This is approximately in line with the data. Using Flow of Funds data and the information in Ueda and Brooks (2011), we find that in 2006 total credit market debt of non-financial businesses with maturities of more than one year reached around 60% of GDP, while residential mortgages[33] reached around 80% of GDP.

Using the parameter s, we calibrate the interest rate margin between the policy rate i_t and the deposit rate $i_{d,t}$ at 1%, consistent with the evidence in Ashcraft and Steindel (2008) and the discussion in Benes and Kumhof (2012). The steady state interest rate margin between the wholesale rate which banks would charge on riskless private loans, $i_{\ell,t}$, and the policy rate, i_t, is fixed at 0.5% per annum in steady state, using the parameter χ. This is roughly equal to the historical spread of the 3-month US$ LIBOR over the 3-month treasury bill rate. The value of the parameter χ implies that in steady state penalty costs equal around 7.7% of the value of assets of those banks that violate the minimum capital adequacy requirement.

For consistency with the specifications of Bernanke et al. (1999) and Christiano et al. (2014), but without loss of generality, we set the steady state value of the willingness-to-lend parameter $\bar{\kappa}$ to 1. For our simulation experiments, the first-order autoregressive coefficients of borrower riskiness shocks and willingness-to-lend shocks are set to 0.9.

For ILF Model 1, we set $\varpi = 0.5$, so that the population sizes of savers and borrowers are equal. The weights on labour disutility ψ^s and ψ^b are calibrated to normalise both groups' steady state labour supply to 1, and the coefficient ι, which determines the sharing of lump-sum income, is calibrated to insure that both groups' steady state per capita consumption levels are equal.

For ILF Model 2, the parameter δ^k is used to calibrate the leverage ratio of entrepreneurs at 100%. The parameter A^c is then set at the same level as in FMC Model 1, which fixes the overall size of the banking sector balance sheet at exactly the same level as in ILF Model 1 and FMC Model 1.

[32] For the core U.S. manufacturing and services sectors the leverage ratio of listed companies has fluctuated around 110% since the early 1990s, but leverage for unlisted companies is likely to have been lower on average due to more constrained access to external financing.

[33] In the data investment in fixed capital also includes residential housing investment.

For FMC Model 2, the elasticity of the physical return to capital $r_{h,t}$ with respect to the amount of capital $\check{k}^h_{t-1}$ is set at $\epsilon_h = 0.001$. At this value the post-shock deviations of user costs $r_{h,t}$ and $r_{k,t}$ from their steady state values are similar in magnitude.

V. Model Impulse Responses to Financial Shocks

Each of the following four sets of illustrative impulse responses, in Figures 5 through 8, is designed to highlight the fundamental differences between ILF and FMC models of banking. The impulse responses are for either credit boom or credit crash scenarios, caused by either borrower riskiness shocks or willingness-to-lend shocks.

The empirical literature has found (Covas and den Haan (2011)) that non-bank corporations tend to have greater difficulty in accessing equity markets during downturns. We will therefore use ILF Model 2 and FMC Model 2 to illustrate credit crashes, because in these models borrowers' access to additional equity is more difficult than in ILF Model 1 and FMC Model 1, which we will use to illustrate credit booms. The reason is that in ILF Model 2 and FMC Model 2 equity is accumulated by entrepreneurs, and can only increase through retained earnings, while in ILF Model 1 and FMC Model 1 equity is accumulated by households, and can in addition increase through private saving decisions.

The empirical literature (Claessens, Kose and Terrones (2011a,b)) has also established that credit booms and crashes are asymmetric, in that gradual prolonged booms in lending have gradual positive effects on economic activity, while crashes are sudden, and are accompanied by sharp contractions in output. Our scenarios account for this, with credit booms being modeled as successive positive shocks over a period of six quarters, and credit crashes as a single large negative shock.

A. Borrower Riskiness Shocks

1. Credit Boom

Figure 5, which is based on ILF Model 1 and FMC Model 1, shows impulse responses for a sequence of unanticipated shocks whereby, over a period of six quarters, the standard deviation of borrower riskiness, σ^k_t, drops by around 30%.

Banks' profitability immediately following each shock is significantly improved at their existing balance sheet and pricing structure. They therefore respond through a combination of lower lending spreads and higher lending volumes. This has a stimulative effect on investment and, through positive wealth effects, on consumption. Because of the increase in aggregate demand, inflation ultimately increases by around 0.7 percentage points at the end of the second year, accompanied by an increase in the real policy rate of around 1 percentage point. Inflation and the policy rate behave fairly similarly across the ILF and FMC models. But the behaviour of the financial sector variables, and consequently of the components of GDP, differs greatly.

ILF banks, in ILF Model 1, cannot quickly change their lending volume, because in order to lend more to borrowers they first have to wait for savers to make sufficient deposits of goods. Savers are willing to do so, but only gradually over time, by cutting back on their consumption of goods and increasing their labour supply and therefore the production of goods. Figure 5 does not show separate impulse responses for saver and borrower households, but this is indeed how saver households behave in this simulation. In other words, because deposits equal the real savings of saver households, and because their savings are a predetermined variable, deposits and thus loans cannot jump following the shock, they have to increase gradually. Banks therefore mostly respond to the favourable shock by reducing their lending spread very significantly, by around 90 basis points at the end of year 1.

FMC banks, in FMC Model 1, technically face no constraints on increasing their loan volume. Because, at the moment when a new loan is made, the borrower is the depositor, banks' main constraint is economic, it is the expected profitability and risk of lending, which of course improves very significantly with each successive reduction in borrower riskiness. Banks therefore increase their lending volume very significantly, by around 8% towards the end of year 2, compared to less than 2% in ILF Model 1. They fund the additional lending by creating additional deposits, and they do so almost one-for-one, as movements in net worth are comparatively small. Ceteris paribus, the much greater increase in lending in FMC Model 1 leads to higher loan-to-value ratios among banks' borrowers. This in turn implies an increase in lending risk, which partly offsets the decrease in lending risk that is due to the favourable shocks to σ_t^k. As a result, banks in FMC Model 1 only reduce their lending spread by around 30 to 40 basis points over the first two years following the shock, compared to 90 basis points in ILF Model 1. In other words, in the FMC model far more of banks' reaction to borrower riskiness shocks consists of quantity rather than price changes. This is also evident in the behaviour of bank leverage. In ILF Model 1 bank leverage decreases on impact because the positive change in bank net worth following the positive earnings surprise dominates the gradual increase in loans. In FMC Model 1 the opposite is true, leverage increases, because the rapid increase in lending dominates the change in net worth. In other words, in the FMC model bank leverage is procyclical, while in the ILF model it is countercyclical.

As for the effects on the real economy, investment increases by around 7% by year 2 in FMC Model 1, compared to little over 2% in ILF Model 1. The main reason is a much greater reduction in the effective price of investment. This reflects increased creation of purchasing power, or liquidity, by banks, which reduces the cost of buying and selling investment goods. The large difference in the investment response implies a sizeable difference in the output response, which is twice as large in the FMC model compared to the ILF model, peaking at over 1.5%. In other words, the way in which banks are modeled matters a great deal not just for predictions for financial sector behaviour, but also for predictions for the real economy.

2. Credit Crash

Figure 6, which is based on ILF Model 2 and FMC Model 2, shows impulse responses for a shock whereby, in a single quarter, the standard deviation of borrower riskiness, σ_t^k, increases by 15%.

Banks' profitability and net worth immediately following this shock are significantly worse at their existing balance sheet and pricing structure. In order to protect their profitability and their capital adequacy levels, they therefore respond through a combination of higher lending spreads and lower lending volumes. This leads to a significant contraction in investment, which is the main reason for the contraction in GDP. Lower demand leads to lower inflation, which in turn leads to a reduction in the real policy rate.

ILF banks, in ILF Model 2, cannot quickly change their lending volume. Because savings are a predetermined variable, deposits and therefore loans can only decline gradually over time, mainly by households increasing their consumption or reducing their labour supply. Banks therefore continue to lend to borrowers that have become much riskier, and to compensate for this risk they increase their lending spread, by around 200 basis points on impact.

FMC banks, in FMC Model 2, technically face no constraints on reducing their loan volume. When spreads rise, entrepreneurs choose to sell some physical capital to households in exchange for deposits, and then use these deposits to repay outstanding loans. In FMC Model 2 there is therefore a large and discrete drop in the size of banks' balance sheet, of around 5% on impact in a single quarter (with almost no initial change in ILF Model 2), as deposits and loans shrink simultaneously. Because, ceteris paribus, this cutback in lending reduces borrowers' loan-to-value ratios and therefore the riskiness of the remaining loans, banks only increase their lending spread by around 100 basis points on impact, compared to 200 basis points in ILF Model 2. As before, a large part of their response is therefore in the form of quantity changes rather than price changes. This is also evident in the behaviour of bank leverage. In ILF Model 2 leverage increases on impact because the net worth losses following the negative earnings surprise dominate the gradual decrease in loans. In FMC Model 2 leverage drops, because the rapid decrease in lending dominates the change in net worth. Again, in the FMC model bank leverage is procyclical, while in the ILF model it is countercyclical.

As for the effects on the real economy, the contraction in GDP in the FMC model is more than twice as large as in the ILF model, and this is due to the behaviour of both investment and consumption. Investment exhibits a large and highly persistent drop because there is a large and persistent cut in lending, which leads to a large contraction in the provision of purchasing power to the economy, and therefore to a large increase in the effective price of investment.

Consumption increases on impact in the ILF model, while it decreases along with investment in the FMC model. The consumption response in the ILF model is in general ambiguous due to several effects that work in opposite directions. But the main effect favouring the observed increase in consumption in ILF Model 2 is due to monetary transactions costs, and is therefore typically not present in other ILF models in the literature.[34] Namely, the borrower riskiness shock has a strong contractionary effect on investment, and therefore also on the demand for transactions balances related to investment. At the same time, the overall supply of transactions balances remains almost constant on impact, as bank deposits, which are real savings in this model, cannot drop rapidly. This frees up some of the existing transactions balances for consumption transactions, so that the effective price of consumption drops on impact, thereby giving a temporary boost to consumption. By contrast, in FMC Model 2 the supply of transactions balances drops very significantly on impact. This increases transactions costs, which in turn sharply increases the effective prices, and therefore the volumes, of both consumption and investment.

[34] Christiano et al. (2014) estimate a model that shares the basic structure of ILF Model 2, but that has different specifications for the interest rate reaction function, nominal rigidities, and real rigidities, and that most importantly does not model bank deposits as being related to the amount of real activity via transactions costs. For a shock that is similar to that in Figure 6, they find that consumption decreases slightly on impact.

The simultaneous drop in consumption and investment in the FMC model is an added advantage of this specification, because a positive co-movement between consumption and investment is a very robust feature of the data. The critical model ingredients that are responsible for this are the FMC specification for banks and the fact that transactions balances are required for both consumption and investment.

Inflation and the policy rate drop by 20-30 basis points more in the FMC model compared to the ILF model, which further reduces the overall increase in interest rates compared to the ILF model. This outcome is directly related to the difference in the behaviour of the supply of transactions balances, with the large discrete contraction in the FMC model adding to the disinflationary pressures.

B. Willingness-to-Lend Shocks

1. Credit Boom

Figure 7, which is based on ILF Model 1 and FMC Model 1, shows impulse responses for a sequence of unanticipated shocks whereby, over a period of six quarters, the share of the value of capital against which banks are willing to lend, κ_t, increases by 30 percentage points.[35] The main observation is that the impulse responses are qualitatively very similar to those in Figure 5. The magnitudes are larger, some differences between models are more accentuated, but the pattern is the same. However, the interpretation of these shocks is different. Specifically, in Figure 7 the fundamentals of borrowers' business have not changed, rather it is changes in sentiment by the banks themselves that are the main driver of this cycle.

In the FMC model, FMC Model 1, the increase in the volume of loans is now so large (14% at the peak) and so fast that lending risk and therefore spreads do not decline at all over the first year following the shock. The reason is that the beneficial effects on lending risk of the shock to κ_t, which are due to a greater availability to banks of collateral per dollar of loans, are offset by a steep increase in loan-to-value ratios that increases the default risk of borrowers. Therefore, in this scenario 100% of banks' initial response to the willingness-to-lend shock consists of quantity rather than price changes. This is very different from the behaviour of the ILF model, where spreads drop by almost 90 basis points by the end of year 1, while the volume of lending increases by barely 1%.

2. Credit Crash

Figure 8, which is based on ILF Model 2 and FMC Model 2, shows impulse responses for a shock whereby, in a single quarter, the share of the value of assets against which banks are willing to lend, κ_t, decreases by 15 percentage points.

Again, the main observation is that the pattern of the impulse responses is very similar to Figure 6, while the magnitudes are larger. The only notable difference concerns the behaviour of bank leverage, which is almost acyclical for the ILF model, while showing a much larger procyclical

[35] Note that $\kappa_t > 1$ does not imply that loans exceed the value of physical capital, because in steady state, where $\bar{\kappa} = 1$, loans equal 50% of the value of physical capital.

drop for the FMC model. The main reason for this difference is the behaviour of bank net worth. The borrower riskiness shock of Figure 6 surprises banks that have inherited loan contracts that were negotiated under previous, more favourable conditions among their borrowers, so that under this shock banks make significant loan losses. The willingness-to-lend shock on the other hand is forward-looking, it affects new loan contracts, with small impact effects on net worth. This means that bank leverage is almost entirely dominated by bank loans, which initially barely change under the ILF model, while dropping by more than 8% under the FMC model.

Table 1 briefly summarizes the main differences in the simulation properties of ILF and FMC models of banking.

VI. Stylised Facts and Related Empirical Literature

Our simulation exercises have generated three interrelated predictions for differences in the behaviour of financial variables in FMC versus ILF models of banking. First, in FMC models changes in the size of bank balance sheets are not limited by the availability of pre-existing savings, and can therefore exhibit potentially very large jumps, while in ILF models they are gradual and modest in size. Second, bank leverage tends to be procyclical in the FMC model, but countercyclical in the ILF model. Third, in the FMC model quantity rationing of credit plays a very important role relative to price rationing, while in the ILF model price rationing dominates.

We now turn to the stylised facts, for major industrialised economies, that relate to these predictions. Our discussion combines a survey of the related empirical literature with, where appropriate, data and empirical analysis of our own.

A. Large Jumps in Credit and Money

In this subsection we show that credit and money exhibit large jumps over the cycle, and also that these jumps are far larger than the contemporaneous changes in national accounts saving, and furthermore sometimes of opposite sign.

In an important paper, Adrian et al. (2013) show that there is very strong positive co-movement between changes in U.S. banks' total assets and total debt (meaning total assets minus net worth), for both aggregate and micro-level data, for both commercial banks and the shadow banking system, and during both booms and recessions. In other words, the U.S. banking system responds to shocks mainly through one-for-one changes in assets and debt, rather than through changes in bank net worth. This includes recessions, when a drop in assets is accompanied mainly by a drop in debt, rather than in net worth.

A temporary exception to this behaviour was the outset of the Great Recession, when the leverage of U.S. commercial banks first increased, with deleveraging only starting in 2009. As discussed in Ivashina and Scharfstein (2009) and Irani (2013), one reason is that large borrowers, for liquidity reasons, initially drew down credit lines that had been approved before the crisis. Another reason is institutional lags - the renegotiation of existing credit lines simply takes time. Lags are therefore clearly important for understanding the behaviour of leverage, and we will return to this in our analysis of the data.

Adrian et al. (2013) illustrate their results using scatter plots, based on FDIC call reports data, that plot dollar changes in debt and equity of a number of U.S. commercial banks against the corresponding dollar changes in assets. Their assets-debt pairs lie along a line with a slope of almost one, while the assets-equity pairs lie along a line with a slope near zero. We repeat a similar exercise in Figure 9, with three differences in the exposition. First, our sample is different, we consider quarter-on-quarter balance sheet changes of the 200 largest U.S. commercial banks by asset size, in the single quarter 2009Q4, a turbulent period with large balance sheet changes due to the financial crisis. Second, we study percent changes in addition to dollar changes. One reason is that using dollar changes gives rise to outliers for very large institutions, which then dominate the assets-debt relationship. Furthermore, one of our concerns is with the behaviour of bank leverage, which depends on percentage changes in assets relative to percentage changes in equity. Third, we also examine histograms that show the distributions of the elasticities of debt and equity with respect to assets.

The exposition in terms of dollar changes in the top left panel of Figure 9 shows a very similar pattern to Adrian et al. (2013). The top right panel shows that for percentage changes the assets-debt pairs continue to lie very close to the 45-degree line, while the assets-equity pairs continue to be widely dispersed. The histograms show that the elasticity of debt with respect to assets is clearly centered at 1, with few outliers. On the other hand, the median and mean elasticities of equity with respect to assets are far lower, and their distribution is far more widely dispersed. Finally, and this is especially important for our argument, for a large number of institutions the percentage changes in assets and debt are very large, approaching 10% or more in this single quarter. It can be shown that in less turbulent quarters the magnitudes are somewhat smaller, but the pattern is the same.

Figure 10 turns to flow-of-funds time series data for the aggregate banking systems of six major industrialised countries (in the U.S. case the aggregate banking system includes commercial banks, shadow banks and government-sponsored enterprises (GSEs)). The pattern is very similar to Figure 9, with the elasticity of debt with respect to assets very close to one across all countries[36], the elasticity of equity with respect to assets far lower and far more widely dispersed, and not infrequent changes in assets and debt of 2%-4% in a single quarter, and in some cases significantly more than that. This shows that banking systems *as a whole* frequently expand or contract assets and debt by very large amounts. We have found that percentage changes in credit are of a similar magnitude to percentage changes in total assets.

The relationship between saving rates and changes in bank debt is taken up in Figure 11. The figure plots two ratios. The first is the change in aggregate financial system adjusted debt in a quarter relative to the GDP of that quarter, where financial system adjusted debt is defined as total assets minus the sum of net worth, interbank deposits, and remaining liabilities (the latter represents mainly derivatives). The second ratio is the net private saving-to-GDP ratio, in other words the ratio of net private saving in a quarter relative to the GDP of that quarter. Saving net of depreciation is the more appropriate concept in a size comparison with changes in bank balance sheets, because the depreciation applicable to an investment project occurs along with the repayment of the underlying debt out of cash flow.

We begin with a detailed discussion of the United States. In the years prior to the onset of the

[36] To the extent that assets and debt change by approximately equal dollar amounts at most times, and given that debt is smaller than assets for solvent institutions, it is expected that the debt-assets regression line has a slope slightly greater than 1.

Great Recession, net private saving fluctuated in a relatively narrow range of 3% to 6% of GDP, while changes in bank liabilities were much larger on average, as the size of the financial system grew relative to the economy, and fluctuated in a far wider range of 6% to 16% of GDP. The financial crisis itself exhibited an extremely large *decline* in the change of bank liabilities relative to GDP, from +14% in early 2008 to -10% in early 2009[37], that was accompanied by a fairly sizeable *increase* in the private saving-to-GDP ratio, from 4% to 9% over the same time period. The behaviour of bank balance sheets was therefore clearly disconnected from that of saving.

Figure 11 shows that other economies exhibit even more extreme patterns than the United States, with bank liabilities frequently changing by over 20% of quarterly GDP (5% of annual GDP) in the Eurozone, Germany and France. All three regions experienced a massive contraction in bank balance sheets at the onset of the Great Recession, with much smaller or no movements in net private saving. Again, the quantity of saving and the size of bank balance sheets is clearly disconnected. Furthermore, the fluctuations in bank balance sheets can be extremely fast and large. This is consistent with the pattern exhibited by the FMC models in Figures 5-8.

Adrian et al. (2013) argue that, for the United States, a switch of corporate borrowers from bank financing to bond financing played an important role in explaining the decline in the size of the financial sector's balance sheet at the beginning of the Great Recession. In other words, the argument is that shrinking bank balance sheets could partly be due to substitution among different forms of financing, rather than to a large absolute decline of financing. However, the empirical evidence for this argument is not very strong. First, we note that for the non-U.S. economies shown in Figure 11, which at the time of the financial crisis display an even more volatile behaviour of bank balance sheets than the United States, a large-scale switch to bond financing is an unlikely explanation, given the smaller sizes of the respective domestic bond markets. Second, for the United States, the top left panel of Figure 11 shows a third time series that uses the same data on corporate bond financing as Adrian et al. (2013). Comparing this time series with that of changes in bank debt[38] demonstrates that, while clearly present, substitution towards corporate bond financing made a fairly limited contribution to explaining the overall decline in the size of bank balance sheets.

The reasons are studied in more detail in Figure 12, which considers changes in bank lending, and which is again based on the same data as Adrian et al. (2013). We observe once more that the increase in bond financing around the onset of the Great Recession was quantitatively comparatively small. The decline in total bank financing, on the other hand, was extremely large. A very large share of this decline (well over 50%) was due to lower household credit, and households do not have the option of substituting towards bond financing. The same is true for non-financial non-corporate businesses, who also experienced a sizeable reduction in bank credit. But even for the non-financial corporate sector, the decline in bank financing far exceeded the increase in bond financing. The dominant aspect of this story is therefore the decline in total financing due to a large decline in the size of bank balance sheets, and not substitution towards non-bank financing.

[37] This contraction of U.S. balance sheets mostly affected shadow banks, while the size of commercial banks' balance sheets did not change dramatically. However, as shown by Chrétien and Lyonnet (2014), commercial banks purchased a large share of the securities previously held by shadow banks by cutting back on other forms of credit. *These* cutbacks in credit are therefore the ultimate reason for the contraction in the balance sheet of the overall banking system.

[38] The relevant comparison for bond financing is with bank lending rather than bank debt. We will turn to this next. However, as shown in Figure 10, bank lending and bank debt move together very closely at all times.

B. Procyclical Bank Leverage

Nuño and Thomas (2013) study the co-movement between the cyclical components of U.S. bank leverage and aggregate output. They show that significant heterogeneity exists across U.S. banks, with commercial banks, savings institutions and credit unions exhibiting acyclical (or sometimes countercyclical) leverage, while shadow banks exhibit strongly procyclical leverage at frequencies of six to 32 quarters.

In our own empirical analysis, which is reported in Tables 2 and 3, we follow Nuño and Thomas (2013) in that we concentrate on co-movements of the cyclical components of bank leverage and output. But we also follow the lead of Ivashina and Scharfstein (2009) in that we take account of lags of output in this relationship. We also extend the analysis beyond the United States to several other large industrialised countries. We find that procyclicality of bank leverage is almost universally observed when output is lagged by one year or more.

To construct leverage ratios we use flow-of-funds data. We define the banking sector's leverage ratio as the ratio of total assets to net worth. We compute the cyclical components of bank leverage and of real GDP by using the Baxter-King (Band-Pass) filter on the logarithms of these two series.

Table 2 shows cross-correlations between the cyclical components of U.S. output and of the leverage ratios of different subsectors of the U.S. financial sector. We observe that leverage in the shadow financial system[39] (excluding GSEs) is strongly procyclical. The correlation with the cyclical component of output is always positive, reaching almost 0.5, and is highly significant for correlations that use output lagged by more than one quarter. The leverage of GSEs is mostly acyclical. Leverage of the regulated banking system is countercyclical with contemporaneous output, and procyclical at output lags of one year or more. This is consistent with our above discussion of institutional lags and drawdowns of pre-existing credit lines in downturns. The behaviour of the aggregate financial system's leverage ratio is similar to that of the regulated banking system, although it is more procyclical at longer lags of output due to the effect of the shadow banking system.[40] It is important to include the shadow banking sector for comparability with the remainder of our analysis, because in non-U.S. data there is no separation between banks and shadow banks.

Table 3 shows that in the Eurozone[41], France and the United Kingdom aggregate leverage is strongly and significantly procyclical at lags of two quarters or more. Germany exhibits a similar pattern, but here procyclicality is weak and not statistically significant. The cyclical behaviour of Japanese and Italian banks' leverage is similar to that in the United States, with countercyclical behaviour contemporaneously, and procyclical behaviour with output lagged by one year or more.

[39] The definition of the shadow financial system is identical to Nuño and Thomas (2013).

[40] Aggregate U.S. data were constructed as the simple, unconsolidated sum of subsectors. Balance sheet positions between different subsectors of the financial system are therefore not netted out. Because this inevitably involves double-counting, caution is required when interpreting our results for the aggregate U.S. financial system.

[41] The sample for Germany ends in 2010Q3 because of a structural break in the data in the following quarter. Specifically the Act Modernizing Accounting of 25 May 2009 mandates that as of December 2010 total bank assets must include trading portfolio derivatives. Our sample for the eurozone, which includes the German data, therefore also ends in 2010Q3.

Current macroeconomic models of banking as a rule do not generate procyclical bank leverage that is driven by the behaviour of loans and deposits.[42] The main reason can be seen clearly in our own ILF Model 2, where the quantitatively most important impact effect of adverse financial shocks is reduced bank net worth, while loans, due to the fact that they represent slow-moving saving, initially change very little. On the other hand, in FMC models such as FMC Model 2, changes in bank assets and debt are large relative to changes in bank net worth. As a result, leverage is procyclical.

C. Quantity Rationing Versus Price Rationing of Credit

In this subsection we discuss an empirical literature that has demonstrated that a major part of banks' response to turning points in the credit cycle consists of adjustments in the quantity of loans rather than in the price of loans. By contrast, as stressed by Waters (2013a), in ILF-style DSGE models financial shocks mostly lead banks to make price rather than quantity adjustments. We have seen this in our simulations of the ILF models, ILF Models 1 and 2.

Following Waters (2013a), we refer to these quantity effects as quantity rationing of credit, but we note that this term has a somewhat different meaning from the existing literature, which is based on the ILF view of banking.[43] Our argument is not that banks' willingness to *lend out* pre-existing loanable funds may be limited by informational asymmetries, or that non-price lending terms may affect the equilibrium quantity of such lending independently of interest rates. Rather, we argue that banks' willingness to *create* new loanable funds through lending always changes along with the price terms that they are prepared to offer to their borrowers.

For the recent period, the empirical evidence of Lown and Morgan (2006), Thies and Gouldey (2010), Waters (2013a,b) and Gilchrist and Zakrajšek (2012) shows that the quantity of credit is a more important driver of real activity than the price of credit.[44] Lown and Morgan (2006) use the Federal Reserve's Survey of Loan Officers Opinion Survey (SLOOS) series to show that lending standards are significantly correlated with the realised quantity of aggregate lending, that they identify fluctuations in credit supply rather than in credit demand, and that they affect real output. Waters (2013a,b) confirms that quantity rationing, as measured by the SLOOS series, is a significant leading indicator for output, and also for capacity utilisation. Lown and Morgan (2006) and Waters (2013a,b) also find that borrowing costs, measured by interest rate spreads, are a less powerful leading indicator. Gilchrist and Zakrajšek (2012) construct a U.S. corporate credit spread and decompose it into a price rationing component that reflects movements in expected default (similar to the retail lending spread of our model), and a quantity rationing component, referred to as the excess bond premium (EBP), which they interpret as reflecting shifts in the effective supply of funds offered by financial intermediaries. They find that the EBP

[42] As explained by Adrian et al. (2013), this applies to Gertler and Karadi (2011), Gertler and Kiyotaki (2011), and also to the models of Brunnermeier and Sannikov (2009, 2010). Nuño and Thomas (2013) develop a general equilibrium model, in the ILF tradition, to explain procyclical bank leverage. However, their model generates procyclicality almost entirely due to the behavior of bank equity, which is not consistent with the facts presented by Adrian et al. (2013).

[43] For early theoretical contributions on quantity rationing, see Jaffee and Russell (1976), Stiglitz and Weiss (1981), Blinder and Stiglitz (1983) and Fuerst (1992). See Bellier, Sayeh and Serve (2012) for a literature survey.

[44] Recent empirical studies using loan-level or bank-level data have also found evidence for the importance of credit rationing. See Ivashina and Scharfstein (2009), Khwaja and Mian (2008) and Kapan and Minoiu (2013). For the Great Depression, Bernanke (1983), Hamilton (1987) and Baum and Thies (1989) also uncover evidence for quantity rationing of credit.

and the SLOOS series, both indicators of quantity rationing, are highly correlated, and that all of the forecasting ability of their spread for the 1985-2010 period is due to the EBP, in other words to changes in quantity rationing of credit.

Figure 13 shows the U.S. output gap (computed by the Congressional Budget Office) between 2000 and 2012, along with two indicators of quantity rationing (an IMF SLOOS series[45] and the Gilchrist and Zakrajšek (2012) EBP), one indicator of price rationing (the spread between Baa and Aaa corporate bond yields), and the real credit gap (the cyclical component of real credit).

We observe that both the SLOOS and the EBP signalled the onset of the crisis early, while the Baa-Aaa spread only started to rise much later. The real credit gap lags the early indicators by several quarters, due to the above-mentioned institutional lags and drawdowns of pre-existing credit lines. In other words, indicators of quantity rationing of credit, but much less so the indicator of price rationing, were powerful early indicators of the subsequent collapse in output, and later in credit.

During the subsequent period the output gap remained large and negative, with only a weak tendency to close. After the immediate crisis period, the Baa-Aaa spread quickly returned to near pre-crisis levels. By contrast, the SLOOS series shows no strong loosening of credit standards to offset the earlier severe tightening, which indicates that the level of quantity rationing remained high. This is also reflected in the detrended measure of credit, which exhibits a similar pattern to the output gap, at least until 2012. If one looks among financial variables for explanations for the prolonged underperformance of the economy, it is therefore quantities rather than prices of credit that hold more promise. This is consistent with the findings of Waters (2013a).

There is therefore strong empirical evidence for the importance of credit rationing at the outset of and throughout financial crises. Macroeconomic models of banking should reflect this fact. Our model simulations in Figures 6 and 8 have shown that in ILF models banks' reaction to adverse financial shocks consists almost entirely of changes in lending spreads, while changes in credit are slow and small. By contrast, in FMC models immediate, large and persistent cutbacks in credit play a very important role.

VII. Conclusions

Economic models that integrate banking with macroeconomics are clearly of the greatest practical relevance at the present time. The currently dominant intermediation of loanable funds (ILF) model views banks as barter institutions that intermediate deposits of pre-existing real loanable funds between depositors and borrowers. The problem with this view is that, in the real world, there are no pre-existing loanable funds, and ILF-type institutions do not exist. Instead, banks create new funds in the act of lending, through matching loan and deposit entries, both in the name of the same customer, on their balance sheets. The financing through money creation (FMC) model reflects this, and therefore views banks as fundamentally monetary institutions. The FMC model also recognises that, in the real world, there is no deposit multiplier mechanism

[45] The individual SLOOS series report, separately for mortgages and corporate loans, the difference between the share (in percentage points) of banks reporting that credit standards have been tightened and the share (in percentage points) of banks reporting that they have been eased in a given quarter. Figure 13 shows an IMF-computed average series over these components of the SLOOS.

that imposes quantitative constraints on banks' ability to create money in this fashion. The main constraint is banks' expectations concerning their profitability and solvency.

In this paper, we have developed and studied simple, illustrative models that reflect the FMC function of banks, and compared them to ILF models. Following identical shocks, FMC models predict changes in bank lending that are far larger, happen much faster, and have much larger effects on the real economy than otherwise identical ILF models, while the adjustment process depends much less on changes in lending spreads. As a result, FMC models are more consistent with several aspects of the data, including large jumps in lending and money, procyclical bank leverage, and quantity rationing of credit during downturns.

Our results suggest that a quantitative investigation of the effects of macroprudential policies using FMC models is likely to yield results that differ significantly from the existing literature. We are confident that this will generate a very useful research agenda.

References

Adrian, T. and Boyarchenko, N. (2013), "Intermediary Leverage Cycles and Financial Stability", Federal Reserve Bank of New York Staff Reports, No. 567.

Adrian, T., Colla, P. and Shin, H.S. (2013), "Which Financial Frictions? Parsing the Evidence from the Financial Crisis of 2007 to 2009", *NBER Macroeconomics Annual*, **27**, 159-214.

Altunbas, Y., Gambacorta, L. and Marques-Ibanez, D. (2009), "Securitization and the Bank Lending Channel", *European Economic Review*, **53**, 996–1009.

Ashcraft, A.B. and Steindel, C. (2008), "Measuring the Impact of Securitization on Imputed Bank Output", Working Paper, Federal Reserve Bank of New York.

Ball, L. (2001), "Another Look at Long-Run Money Demand", *Journal of Monetary Economcics*, **47**, 31-44.

Baum, C. and Thies, C. (1989), "The Term Structure of Interest Rates and the Demand for Money During the Great Depression", *Southern Economic Journal*, **56(2)**, 490-498.

Bellier, A., Sayeh, W. and Serve, S. (2012), "What Lies Behind Credit Rationing? A Survey of the Literature", Université de Cergy Pontoise, Thema Working Papers, No. 2012-39.

Benes, J. and Kumhof, M. (2012), "The Chicago Plan Revisited", IMF Working Paper WP/12/202.

Bernanke, B. (1983), "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression", *American Economic Review*, **73(3)**, 257-276.

Bernanke, B., Gertler, M. and Gilchrist, S. (1999), "The Financial Accelerator in a Quantitative Business Cycle Framework", in: Taylor, J.B., Woodford, M. (eds.), *Handbook of Macroeconomics, Volume 1C*. Amsterdam: Elsevier, pp. 1341-1393.

Berry, S., Harrison, R., Thomas, R. and de Weymarn, I. (2007), "Interpreting Movements in Broad Money", *Bank of England Quarterly Bulletin*, 2007 Q3, 47(3), 376-88.

Bindseil, U. (2004), "The Operational Target of Monetary Policy and the Rise and Fall of Reserve Position Doctrine", European Central Bank Working Paper Series, No. 372.

Blinder, A. and Stiglitz, J. (1983), "Money, Credit Constraints and Economic Activity", *American Economic Review, Papers and Proceedings*, **73**, 297-302.

Boissay, F., Collard, F. and Smets, F. (2013), "Booms and Banking Crises", Working Paper.

Borio, C. and Disyatat, P. (2009), "Unconventional Monetary Policy: An Appraisal", BIS Working Papers, No. 292.

Borio, C. and Disyatat, P. (2011), "Global Imbalances and the Financial Crisis: Link or No Link?", BIS Working Papers, No. 346.

Brunner, K. and Meltzer, A. (1990), "Money Supply", in: Friedman, B. and Hahn, F. (eds.), *Handbook of Monetary Economics*, Amsterdam: North Holland, 357-398.

Brunnermeier, M. and Sannikov, Y. (2009), "A Macroeconomic Model with a Financial Sector", Working Paper, Princeton University.

Brunnermeier, M. and Sannikov, Y. (2010), "The i Theory of Money", Working Paper, Princeton University.

Bundesbank (2012), *Geld und Geldpolitik*, Frankfurt am Main: Deutsche Bundesbank.

Calvo, G.A. (1983), "Staggered Prices in a Utility-Maximizing Framework", *Journal of Monetary Economics*, **12**, 383-398.

Carpenter, S. and Demiralp, S. (2010), "Money, Reserves, and the Transmission of Monetary

Policy: Does the Money Multiplier Exist?”, Finance and Economics Discussion Series, No. 2010-41, Divisions of Research & Statistics and Monetary Affairs, Washington, DC: Federal Reserve Board.

Carvalho, V., Martin, A. and Ventura, J. (2012), “Understanding Bubbly Episodes”, *American Economic Review, Papers and Proceedings*, 1-6.

Chari, V.V. and Phelan, C. (2014), “On the Social Usefulness of Fractional Reserve Banking”, *Journal of Monetary Economics*, **65**, 1-13.

Chrétien, E. and Lyonnet, V. (2014), “Fire Sales and the Banking System”, Working Paper, École Polytechnique, Paris.

Christiano, L. J., Eichenbaum, M. and Evans, C. L. (2005), “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy,” *Journal of Political Economy*, **113(1)**, 1–45.

Christiano, L., Motto, R. and Rostagno, M. (2014), “Risk Shocks”, *American Economic Review*, **104(1)**, 27-65.

Claessens, S., Kose, A. and Terrones, M. (2011a), “Financial Cycles: What? How? When?”, IMF Working Papers, WP/11/76.

Claessens, S., Kose, A. and Terrones, M. (2011b), “How Do Business and Financial Cycles Interact?”, IMF Working Papers, WP/11/88.

Clerc, L., Derviz, A., Mendicino, C., Moyen, S., Nikolov, K., Stracca, L., Suarez, J. and Vardoulakis, A. (2014), “Capital Regulation in a Macroeconomic Model with Three Layers of Default”, Working Paper, European System of Central Banks.

Covas, F. and den Haan, W. (2011), “The Cyclical Behavior of Debt and Equity Finance”, *American Economic Review*, **101**, 877-899.

Culbertson, J. (1958), “Intermediaries and Monetary Theory: A Criticism of the Gurley-Shaw Theory”, *American Economic Review*, **48(1)**, 119-131.

Curdia, V. and Woodford, M. (2010), “Credit Spreads and Monetary Policy”, *Journal of Money, Credit and Banking*, **24(6)**, 3–35.

De Fiore, F., Teles, P. and Tristani, O. (2011), “Monetary Policy and the Financing of Firms”, *American Economic Journal: Macroeconomics*, **3(4)**, 1112-1142.

Disyatat, P. (2010), “The Bank Lending Channel Revisited”, Bank for International Settlements, BIS Working Papers No. 297.

Douglas, P. (1935), *Controlling Depressions*, New York: W.W. Norton & Co., Inc.

Dudley, W. (2009), “The Economic Outlook and the Fed’s Balance Sheet: The Issue of "How" versus "When"”, Remarks at the Association for a Better New York Breakfast Meeting, available at http://www.newyorkfed.org/newsevents/speeches/2009/dud090729.html.

European Central Bank (2012), *Monthly Bulletin*, May, Frankfurt: European Central Bank.

Fisher, I. (1935), *100% Money: Designed to keep checking banks 100% liquid; to prevent inflation and deflation; largely to cure or prevent depressions; and to wipe out much of the National Debt*, New York: The Adelphi Company.

Fisher, I. (1936), “100% Money and the Public Debt”, *Economic Forum*, Spring Number, April-June 1936, 406-420.

Friedman, M. (1971), “The Euro-Dollar Market: Some First Principles”, Federal Reserve Bank of St. Louis Review, 16-24.

Fuerst, T. (1992), “Liquidity, Loanable Funds and Real Activity”, *Journal of Monetary*

Economics, **29**, 3-24.

Gerali, A., Neri, S., Sessa, L. and Signoretti, F. (2010), "Credit and Banking in a DSGE Model of the Euro Area", Bank of Italy Working Paper Series, No. 740.

Gertler, M. and Karadi, P. (2011), "A Model of Unconventional Monetary Policy", *Journal of Monetary Economics*, **58(1)**, 17-34.

Gertler, M. and Kiyotaki, N. (2011), "Financial Intermediation and Credit Policy in Business Cycle Analysis", in: Friedman, B. and Woodford, M. (eds.), *Handbook of Monetary Economics, Volume 3A*, North Holland: Elsevier.

Gilchrist, S. and Zakrajšek, E. (2012), "Credit Spreads and Business Cycle Fluctuations", *American Economic Review*, **102(4)**, 1692-1720.

Goodfriend, M. and McCallum, B. (2007), "Banking and Interest Rates in Monetary Policy Analysis: A Quantitative Exploration", *Journal of Monetary Economics*, **54**, 1480-1507.

Goodhart, C. (2007), "Whatever Became of the Monetary Aggregates?", adapted from the Peston Lecture in honour of Maurice, Lord Peston, delivered at Queen Mary College, London, on February 28, 2007.

Goodhart, C., Kashyap, A., Tsomocos, D. and Vardoulakis, A. (2012), "Financial Regulation in General Equilibrium", NBER Working Papers, No. 17909.

Graham, F. (1936), "Partial Reserve Money and the 100 Per Cent Proposal", *American Economic Review*, 428-40.

Graziani, A. (1989), "The Theory of the Monetary Circuit", *Thames Papers in Political Economy*, Spring, 1-26.

Gurley, J. and Shaw, E. (1955), "Financial Aspects of Economic Development", *American Economic Review*, **45**, 515-538.

Gurley, J. and Shaw, E. (1956), "Financial Intermediaries and the Saving-Investment Process", *Journal of Finance,* **11**, 257-276.

Hamilton, J. (1987), "Monetary Factors in the Great Depression", *Journal of Monetary Economics*, **19(2)**, 147-170.

Holmes, A. (1969), "Operational Constraints on the Stabilization of Money Supply Growth", in: *Controlling Monetary Aggregates*, Boston: Federal Reserve Bank of Boston, 65-77.

Hudson, M. (2012), *The Bubble and Beyond. Fictitious Capital, Debt Deflation and Global Crisis*, Dresden, Germany: ISLET-Verlag.

Iacoviello, M. (2005), "House Prices, Borrowing Constraints and Monetary Policy in the Business Cycle", *American Economic Review*, **95(3)**, 739-764.

Irani, R. (2013), "Bank Health and Corporate Liquidity Provision", Working Paper.

Ireland, P. (2001), "Sticky-Price Models of the Business Cycle: Specification and Stability", *Journal of Monetary Economics*, **47**, 3-18.

Ivashina, V. and Scharfstein, D. (2009), "Bank Lending During the Financial Crisis of 2008", Working Paper, Harvard Business School.

Jaffee, D. and Russell, T. (1976), "Imperfect Information and Credit Rationing", *Quarterly Journal of Economics*, **90**, 651-666.

Jermann, U. and Quadrini, V. (2012), "Macroeconomic Effects of Financial Shocks", *American Economic Review*, **102(1)**, 238-271.

Kapan, T. and Minoiu, C. (2013), "Liquidity Shocks and the Supply of Credit After the 2007-08 Crisis", Working Paper, Fannie Mae and International Monetary Fund.

Kashyap, A. and Stein, J. (1993), “Monetary Policy and Bank Lending”, NBER Working Papers, No. 4317.

Kashyap, A., Stein, J. and Wilcox, D. (1993), “Monetary Policy and Credit Conditions: Evidence from the Composition of External Finance”, *American Economic Review*, **83(1)**, 78-98.

Keister, T. and McAndrews, J. (2009), “Why Are Banks Holding So Many Excess Reserves?”, Federal Reserve Bank of New York, Staff Report No. 380.

Khwaja, A. and Mian, A. (2008), “Tracing the Impact of Bank Liquidity Shocks: Evidence from an Emerging Market”, *American Economic Review*, **98(4)**, 1413-1442.

King, M. (2012), Speech to the South Wales Chamber of Commerce at the Millenium Centre, Cardiff, October 23.

Kiyotaki, N. and Moore, J. (1997), “Credit Cycles”, *Journal of Political Economy*, **105(2)**, 211-248.

Knight, F. (1927), “Review of Frederick Soddy’s ’Wealth, Virtual Wealth, and Debt’”, *The Saturday Review of Literature* (April 16), p. 732.

Knight, F. (1933), “Memorandum on Banking Reform”, March, Franklin D. Roosevelt Presidential Library, President’s Personal File 431.

Krugman, P. (2013), “Commercial Banks as Creators of "Money"”, New York Times, August 24, available at http://krugman.blogs.nytimes.com/2013/08/24/commercial-banks-as-creators-of-money/?_r=1.

Kydland, F. and Prescott, E. (1990), “Business Cycles: Real Facts and a Monetary Myth”, Federal Reserve Bank of Minneapolis Quarterly Review, **14(2)**, 3-18.

Lindner, F. (2012), “Saving Does Not Finance Investment: Accounting as an Indispensable Guide to Economic Theory”, IMK Working Papers, No. 100.

Lindner, F. (2013), “Does Saving Increase the Supply of Credit? A Critique of Loanable Funds Theory”, Working Paper, IMK.

Lombra, R. (1992), “Understanding the Remarkable Survival of Multiplier Models of Money Stock Determination”, *Eastern Economic Journal*, **18(3)**, 305-314.

Lown, C. and Morgan, D. (2006), “The Credit Cycle and the Business Cycle: New Findings Using the Loan Officer Opinion Survey”, *Journal of Money, Credit and Banking*, **38(6)**, 1575-1597.

Macmillan Committee (1931), *British Parliamentary Reports on International Finance: The Report of the Macmillan Committee*. London: H.M. Stationery Office.

Martin, A. and Ventura, J. (2012), “Economic Growth with Bubbles”, *American Economic Review*, **102(6)**, 3033-3058.

McLeay, M., Radia, A. and Thomas, R. (2014a), “Money in the Modern Economy: An Introduction”, *Bank of England Quarterly Bulletin*, **54(1)**, 4-13.

McLeay, M., Radia, A. and Thomas, R. (2014b), “Money Creation in the Modern Economy”, *Bank of England Quarterly Bulletin*, **54(1)**, 14-27.

Minsky, H. (1986), “Money and Crisis in Schumpeter and Keynes”, in: Wagener, H.-J., Drukker, J.W. (eds.), *The Economic Law of Motion of Modern Society: A Marx-Keynes-Schumpeter Centennial*. Cambridge, New York and Sydney: Cambridge University Press, pp. 112-122.

Minsky, H. (1991), “Endogeneity of Money”, in: Nell, E.J., Semmler, W. (eds.), *Nicholas Kaldor and Mainstream Economics: Confrontation or Convergence?*, New York, NY: St. Martin’s Press, pp. 207-220.

Moore, B. (1979), “The Endogenous Money Stock”, *Journal of Post Keynesian Economics*, **2(1)**, 49-70.

Moore, B. (1983), “Unpacking the Post Keynesian Black Box: Bank Lending and the Money Supply”, *Journal of Post Keynesian Economics*, **5(4)**, 537-556.

Nuño, G. and Thomas, C. (2013), “Bank Leverage Cycles”, European Central Bank, Working Paper Series, No. 1524.

Pozsar, Z. (2014), “Shadow Banking: The Money View”, OFR Working Papers.

Rogers, J. (1929), *The Process of Inflation in France 1914-1927*, New York: Columbia University Press.

Schmitt-Grohé, S. and Uribe, M. (2004), “Optimal Fiscal and Monetary Policy under Sticky Prices”, *Journal of Economic Theory*, **114**, 198-230.

Schumpeter, J.A. (1954), *History of Economic Analysis*, New York: Oxford University Press.

Simons, H. (1946), “Debt Policy and Banking Policy”, *Review of Economic Statistics*, **28(2)**, 85-89.

Simons, H. (1948), *Economic Policy for a Free Society*, Chicago: University of Chicago Press.

Smith, W. (1959), “Financial Intermediaries and Monetary Controls”, *Quarterly Journal of Economics*, **73,** 533-553.

Standard and Poor’s (2013), “Repeat After Me: Banks Cannot and Do Not "Lend Out" Reserves”, Ratings Direct, Paul Sheard, Chief Global Economist, August 13.

Stiglitz, J. and Weiss, A. (1981), “Credit Rationing in Markets with Imperfect Information”, *American Economic Review*, **71(2)**, 393-410.

Thies, C. and Gouldey, B. (2010), “Direct Measures of Credit Rationing, 2002-2010”, *Journal of the Northeastern Association of Business, Economics and Technology*, 45-51.

Tobin, J. (1963), “Commercial Banks as Creators of Money”, Cowles Foundation Discussion Papers, No. 159.

Towers, G. (1939), *Minutes of Proceedings and Evidence Respecting the Bank of Canada*, Committee on Banking and Commerce, Ottawa: Government Printing Bureau.

Turner, A. (2013), “Credit, Money and Leverage”, September 12.

Ueda, K. and Brooks, R. (2011), “User Manual for the Corporate Vulnerability Utility: The 4th Edition”, International Monetary Fund.

Waters, G. (2013a), “Quantity Rationing of Credit and the Phillips Curve”, *Journal of Macroeconomics*, **37**, 68-80.

Waters, G. (2013b), “Quantity versus Price Rationing of Credit: An Empirical Test”, *International Journal of Financial Studies*, **1**, 45-53.

Werner, R. (2005), *New Paradigm in Macroeconomics: Solving the Riddle of Japanese Macroeconomic Performance*, Basingstoke: Palgrave Macmillan.

Werner, R. (2014a), “Can Banks Individually Create Money Out of Nothing? The Theories and the Empirical Evidence”, *International Review of Financial Analysis*, **36**, 1-19.

Werner, R. (2014b), “How Do Banks Create Money, and Why Can Other Firms Not Do the Same?”, *International Review of Financial Analysis*, **36**, 71-77.

Wicksell, K. (1906), *Lectures on Political Economy, Volume Two: Money*, Lionel Robbins, ed., London: Routledge and Sons, Ltd.

Table 1. Main Simulation Properties of ILF and FMC Models

	ILF	FMC
Loans: Relative Speed of Change	Slower	Faster
Loans: Relative Size of Change	Smaller	Larger
Spreads: Relative Size of Change	Larger	Smaller
Bank Leverage: Cyclicality	Countercyclical	Procyclical
GDP: Relative Size of Change	Smaller	Larger

Table 2. Cross-Correlation of Financial Sector Leverage and Output in the United States

	Regulated	Shadow(ex GSE)	GSE	Total (ex GSE)	Total
	Sample: 88Q1-13Q2				
Lags					
0	-0.35 **	0.12	-0.01	-0.31 **	-0.43 **
1	-0.35 **	0.32 **	-0.10	-0.11	-0.23 **
2	-0.22	0.45 **	-0.13	0.16 *	0.05
3	-0.01	0.49 **	-0.13	0.43 **	0.34 **
4	0.22 **	0.46 **	-0.11	0.62 **	0.56 **
5	0.36 **	0.35 **	-0.11	0.67 **	0.63 **

Cross-correlation between cyclical components of logarithm of lagged GDP and leverage ratio.

Regulated: Depository Institutions. Shadow (ex GSE): Security Brokers/Dealers and Finance Companies.

Total: Unconsolidated sum of subsectors.

Source: U.S. Flow-of-Funds.

* = Significant at 5% confidence level, ** = Significant at 1% confidence level.

Table 3. Cross-Correlation of Financial Sector Leverage and Output in Europe and Japan

	Eurozone	Germany	France	Italy	UK	Japan
Sample:	97Q3-10Q3	80Q1-10Q3	80Q1-13Q2	97Q1-13Q2	98Q2-13Q2	97Q4-13Q2
Lags						
0	0.00	-0.01	-0.01	-0.61 **	0.22	-0.32 *
1	0.29 *	0.09	0.11	-0.47 **	0.51 **	-0.24
2	0.48 **	0.14	0.23 **	-0.25 *	0.70 **	-0.10
3	0.50 **	0.15	0.30 **	0.01	0.76 **	0.05
4	0.39 **	0.11	0.32 **	0.26 *	0.68 **	0.16
5	0.20	0.05	0.27 **	0.45 **	0.48 **	0.22

Cross-correlation between cyclical components of logarithm of lagged GDP and leverage ratio.

Source: See the Data Appendix (available from the authors on request).

* = Significant at 5% confidence level, ** = Significant at 1% confidence level.

Figure 1. ILF Banks: The Naïve Partial Equilibrium View

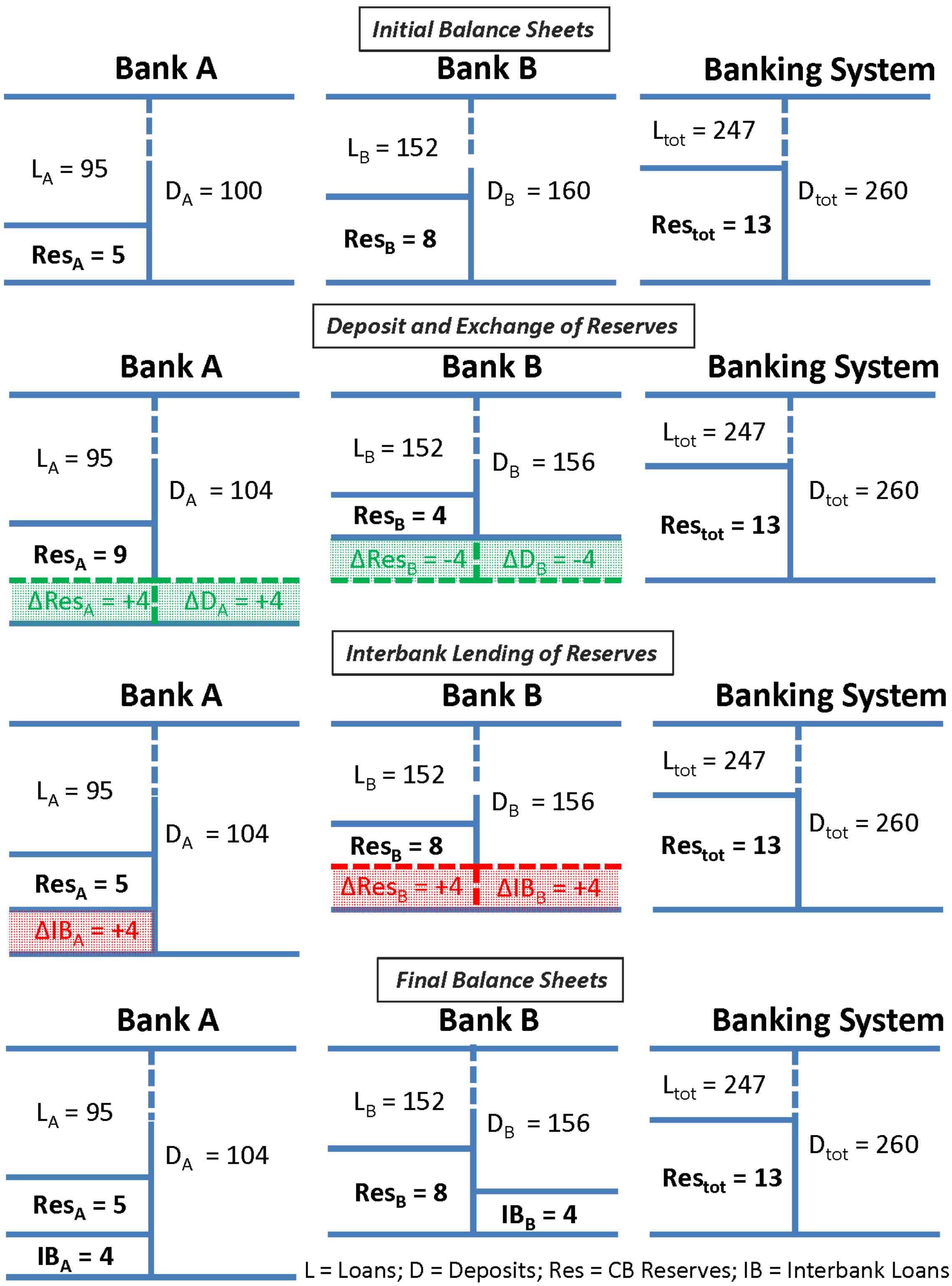


Figure 2. ILF Banks: The Implicit Conventional View

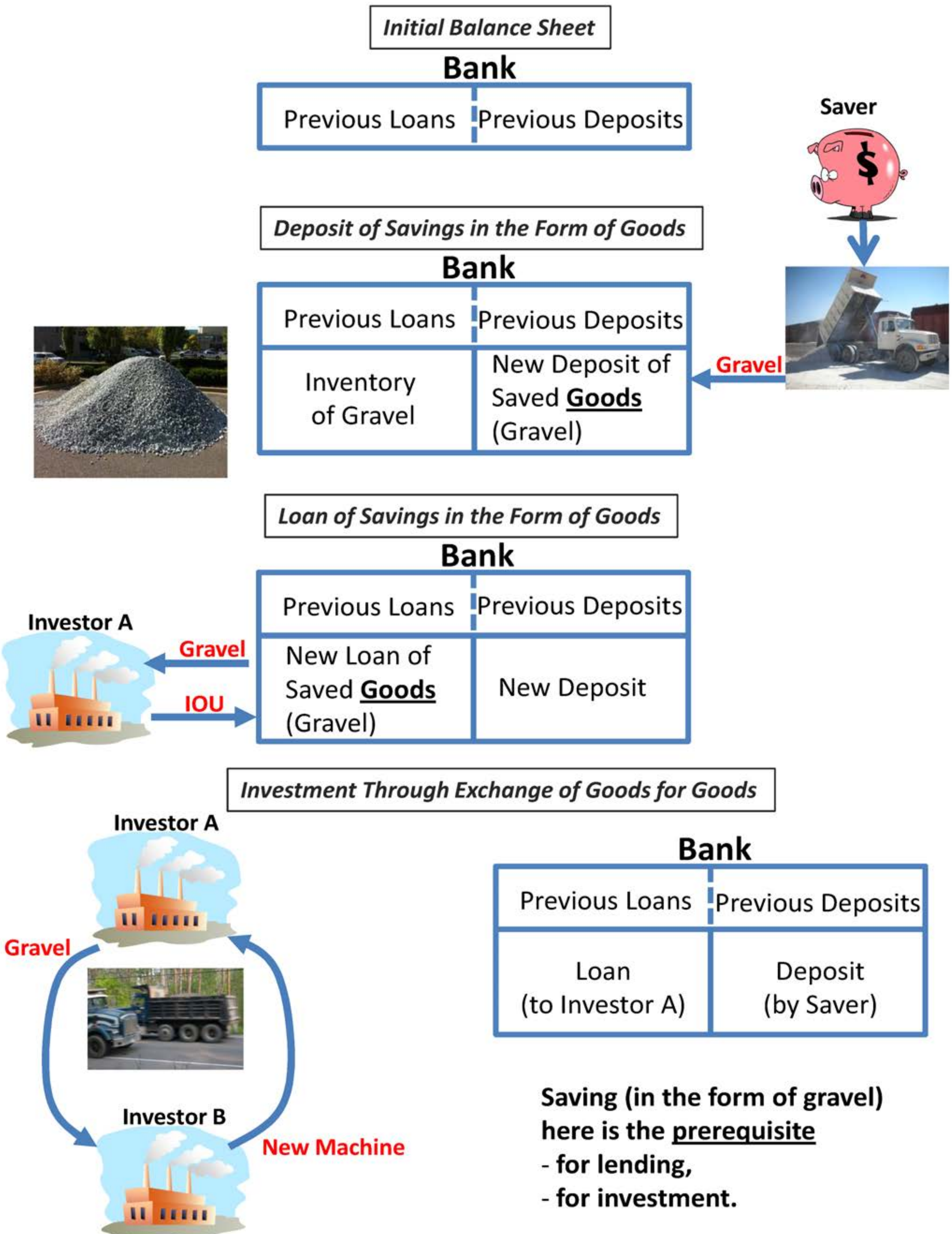


Figure 3. FMC Banks: The Correct View

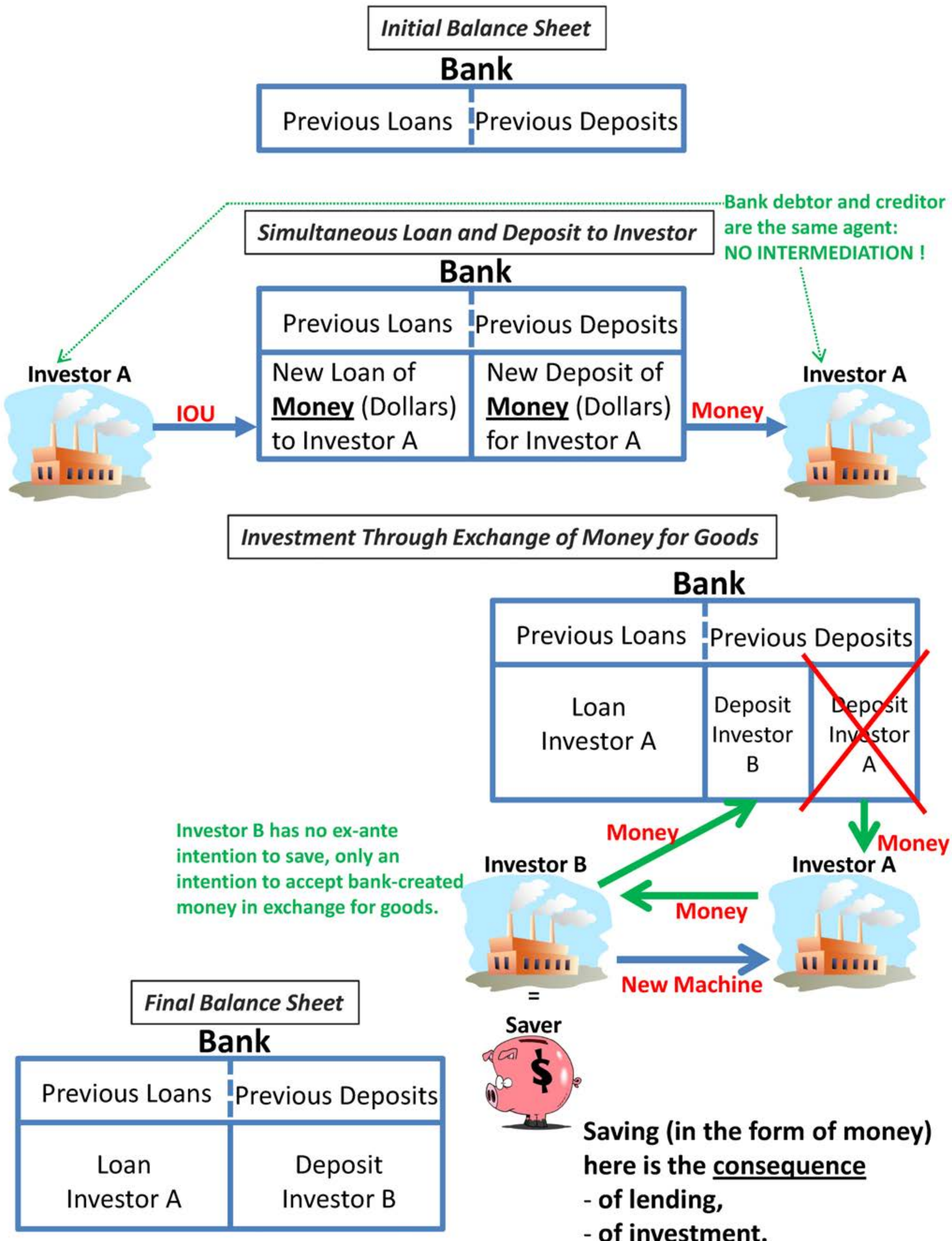


Figure 4. The Role of Banks in the Four Models

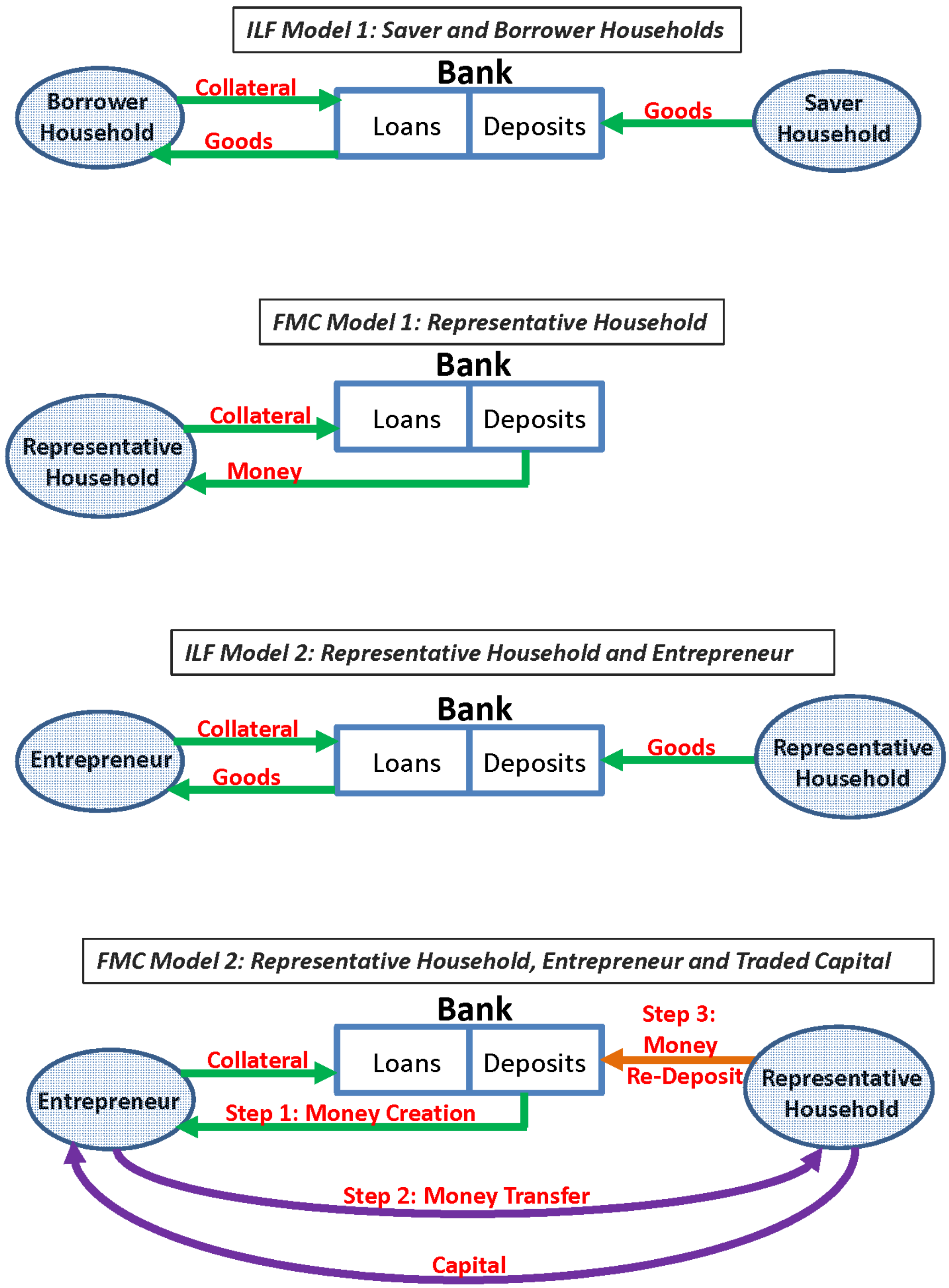
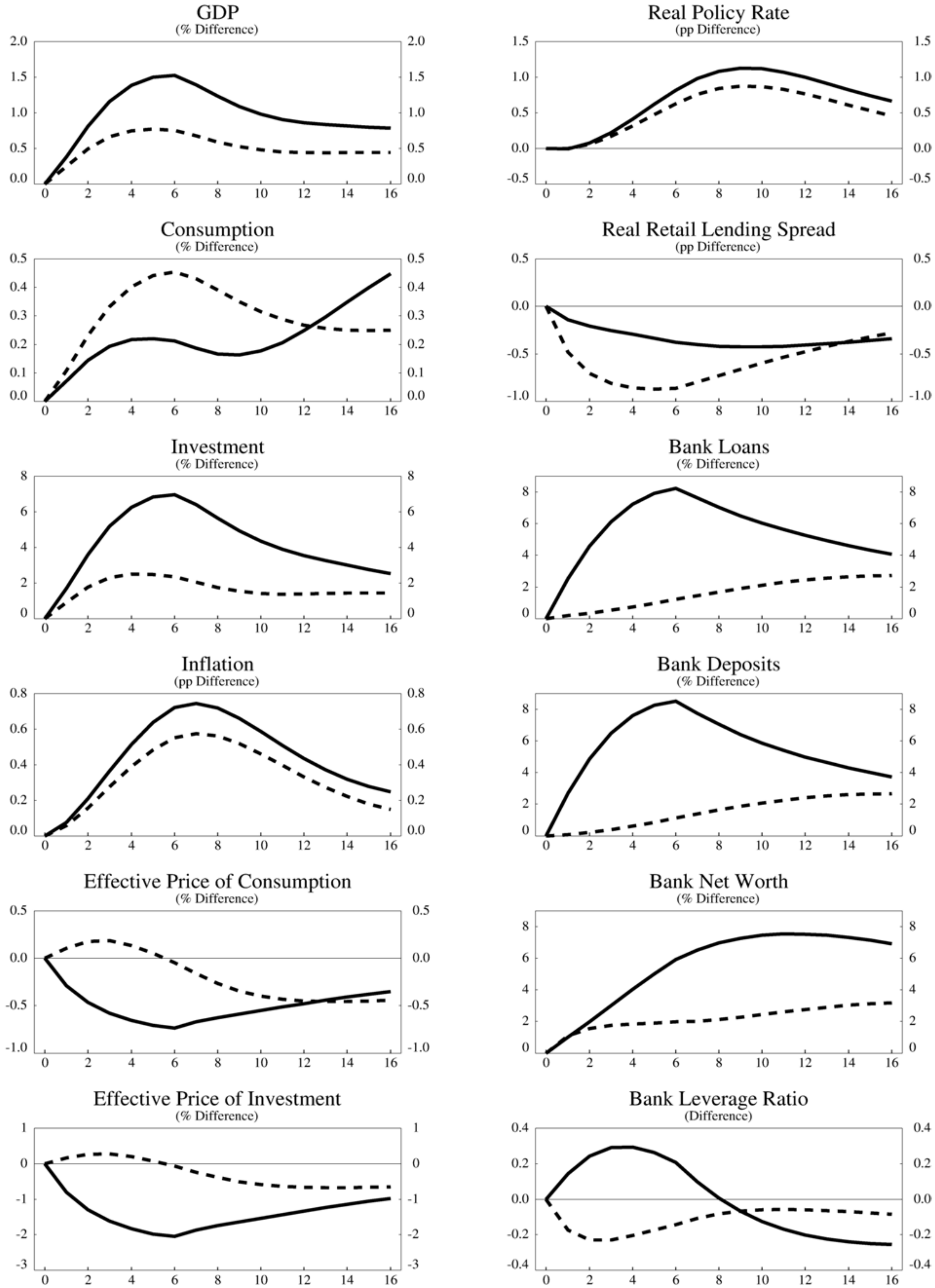
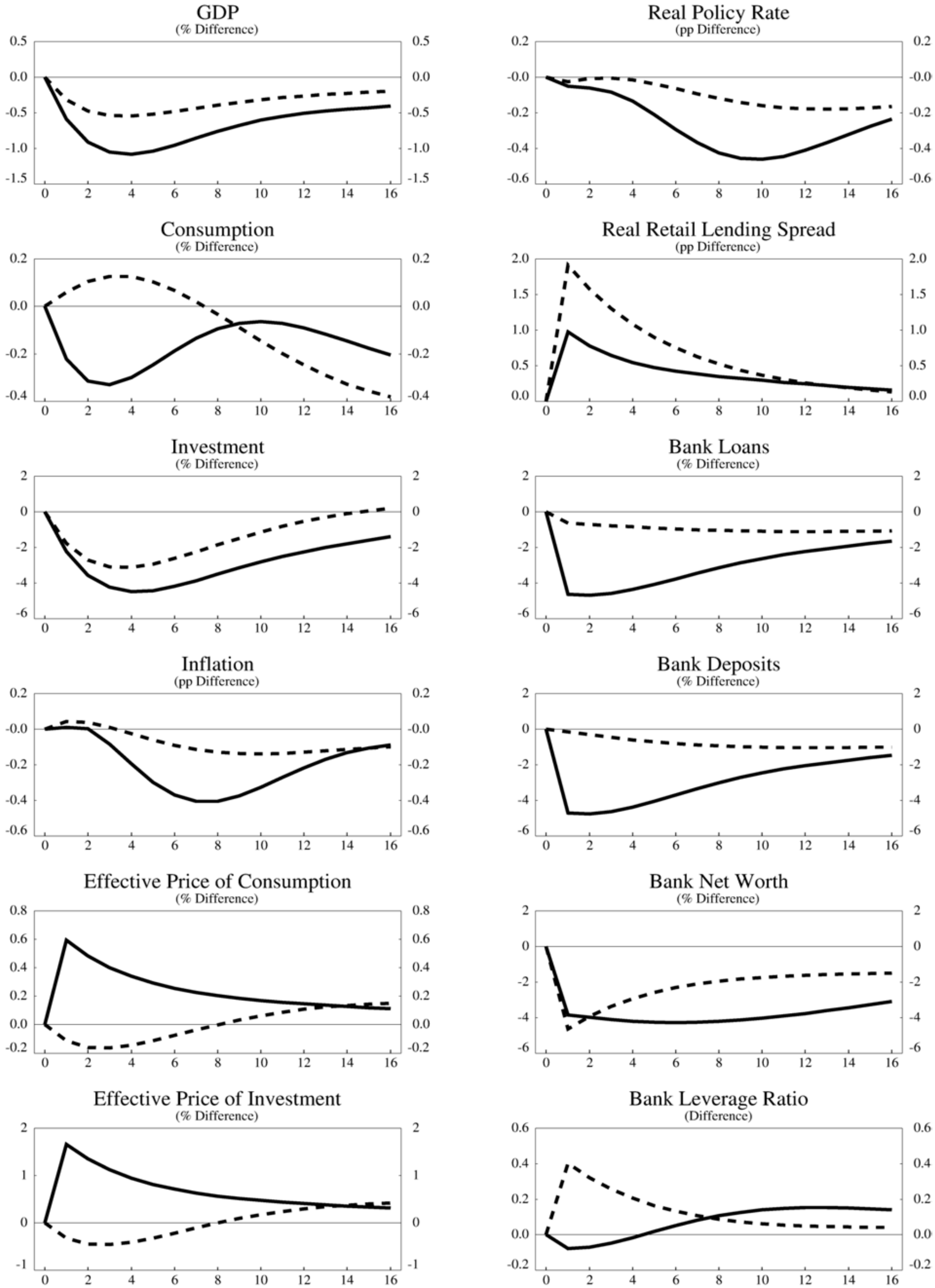


Figure 5. Impulse Responses: Credit Boom due to Lower Borrower Riskiness



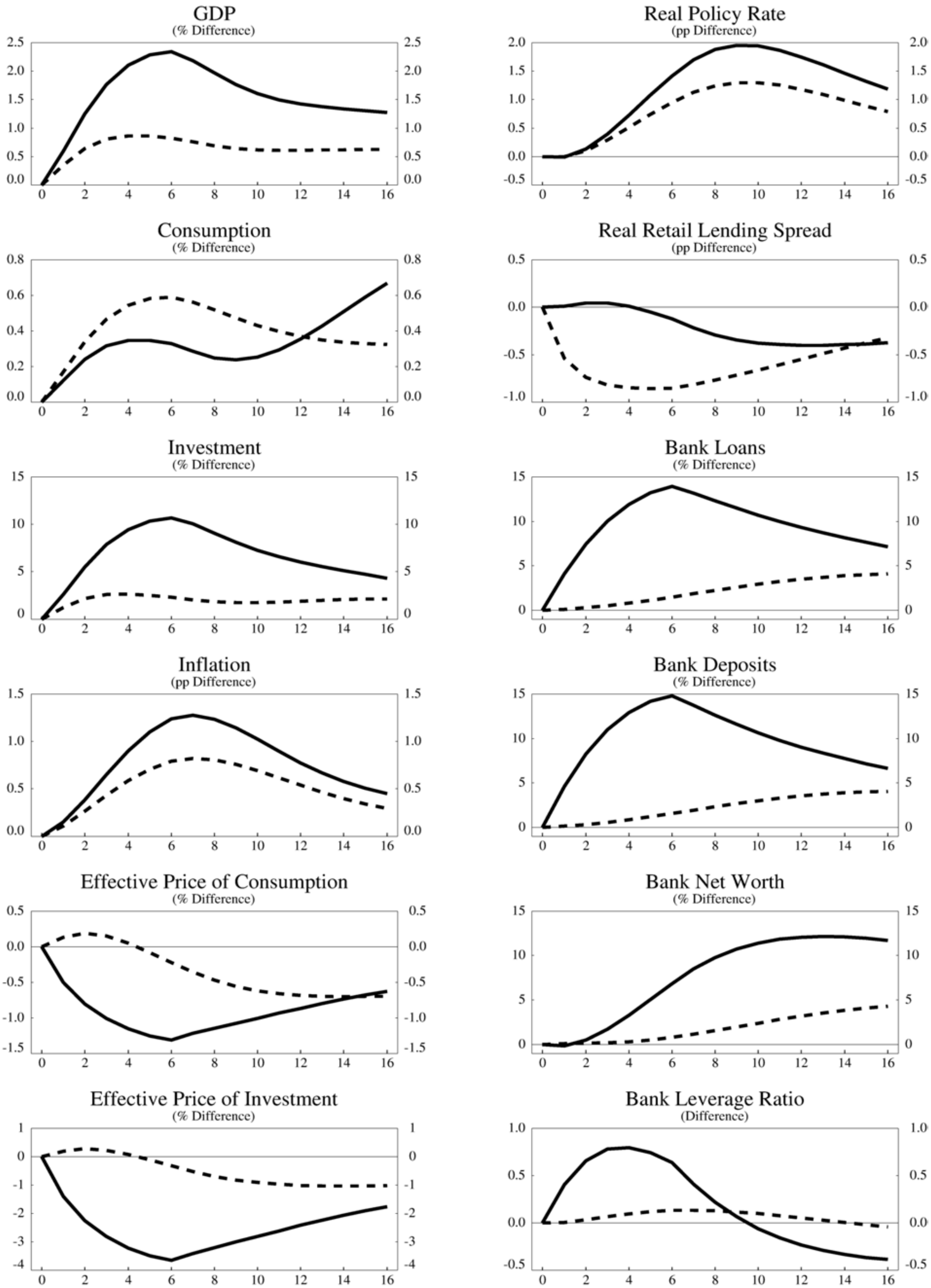
- - - = ILF Model, —— = FMC Model

Figure 6. Impulse Responses: Credit Crash due to Higher Borrower Riskiness



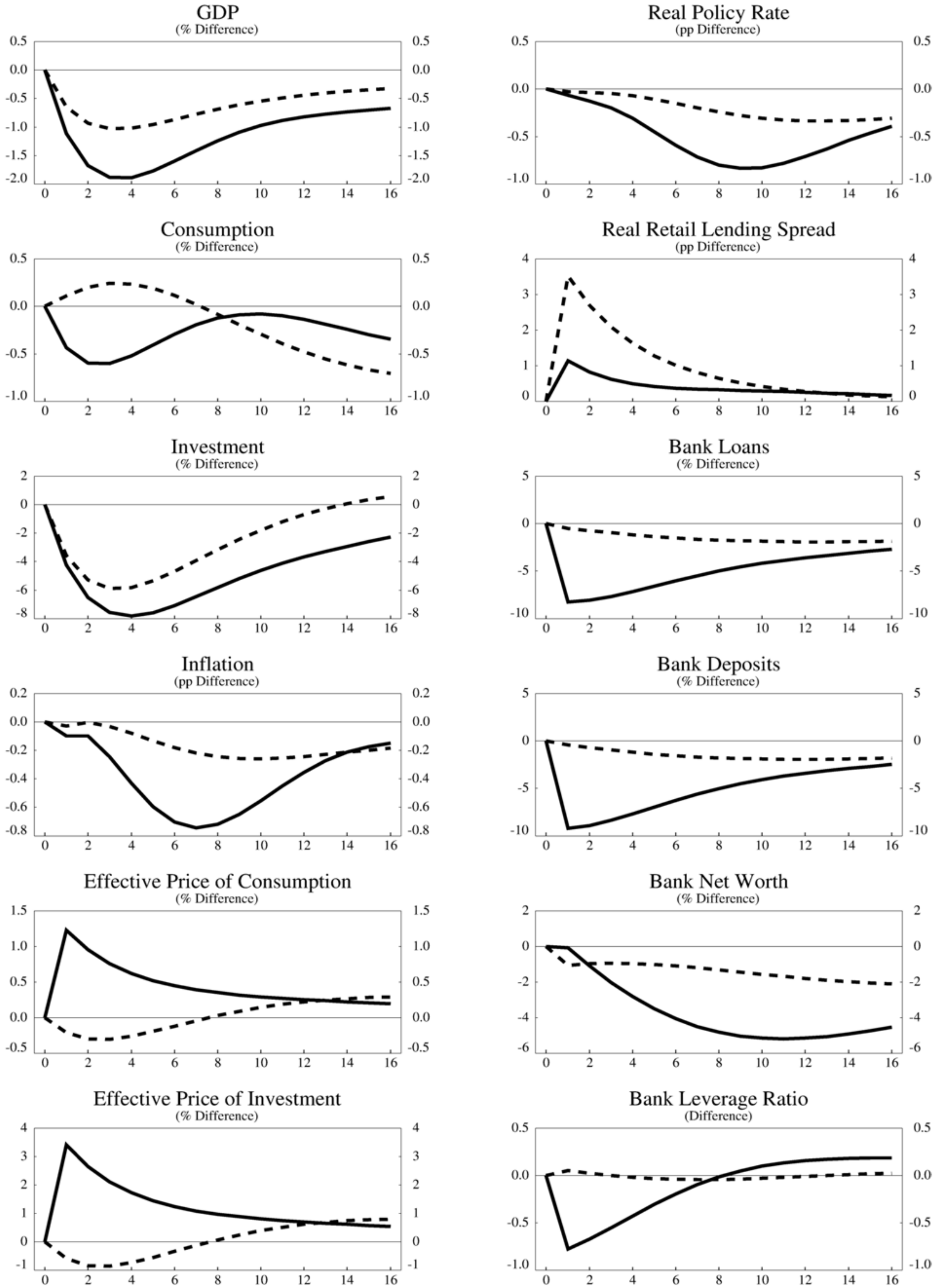
- - - = ILF Model, —— = FMC Model

Figure 7. Impulse Responses: Credit Boom due to Higher Willingness to Lend



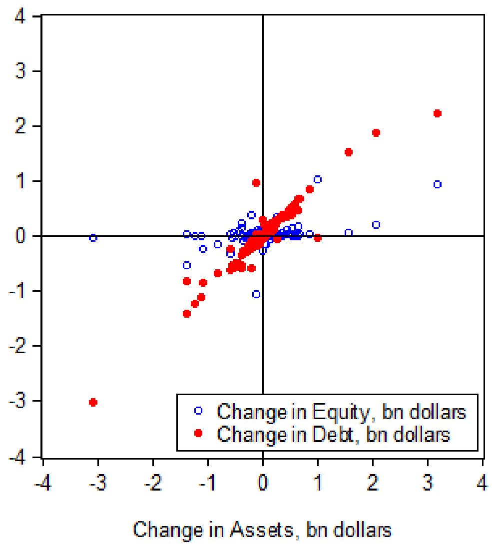
- - - = ILF Model, —— = FMC Model

Figure 8. Impulse Responses: Credit Crash due to Lower Willingness to Lend

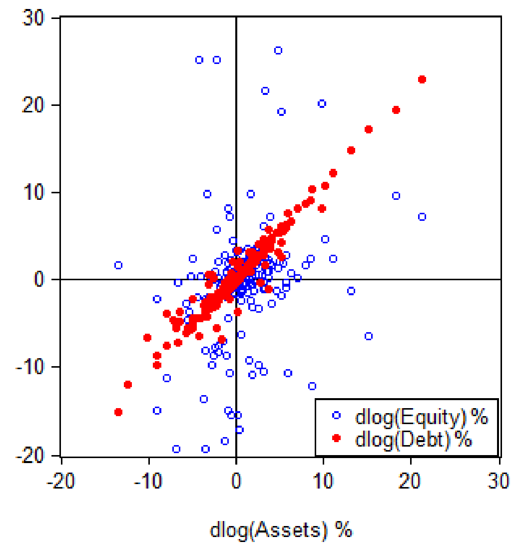


- - - = ILF Model, —— = FMC Model

Figure 9. Bank Balance Sheets: Cross-Sectional Evidence for the United States

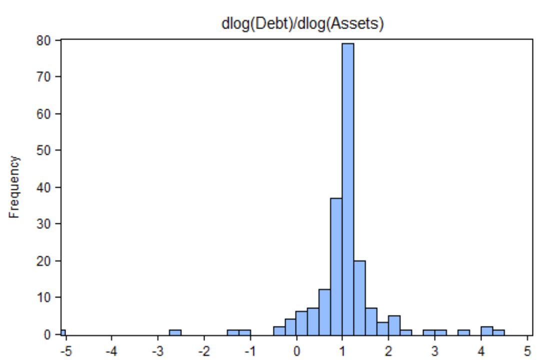


Dollar Changes

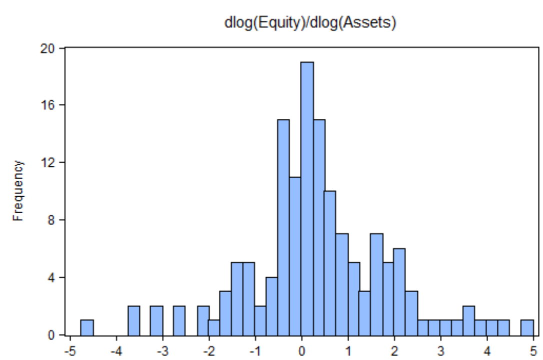


Percent Changes

2009Q4 on 2009Q3 dollar and percent changes in individual banks' assets, debt and equity. Data: FDIC call reports. 200 largest U.S. banks by asset size.

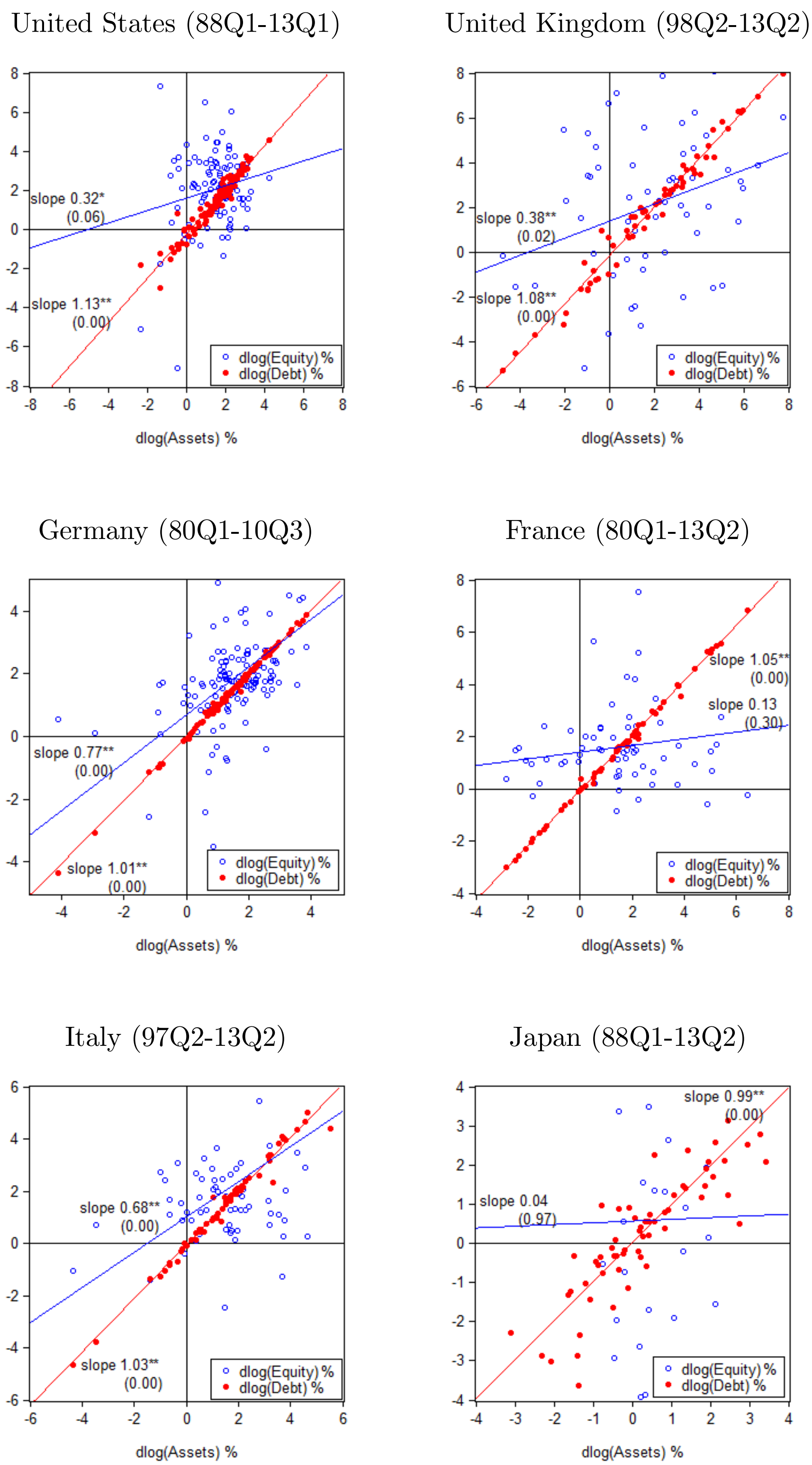


Histogram of 2009Q4 on 2009Q3 elasticity of debt with respect to assets. 200 largest U.S. banks by asset size.



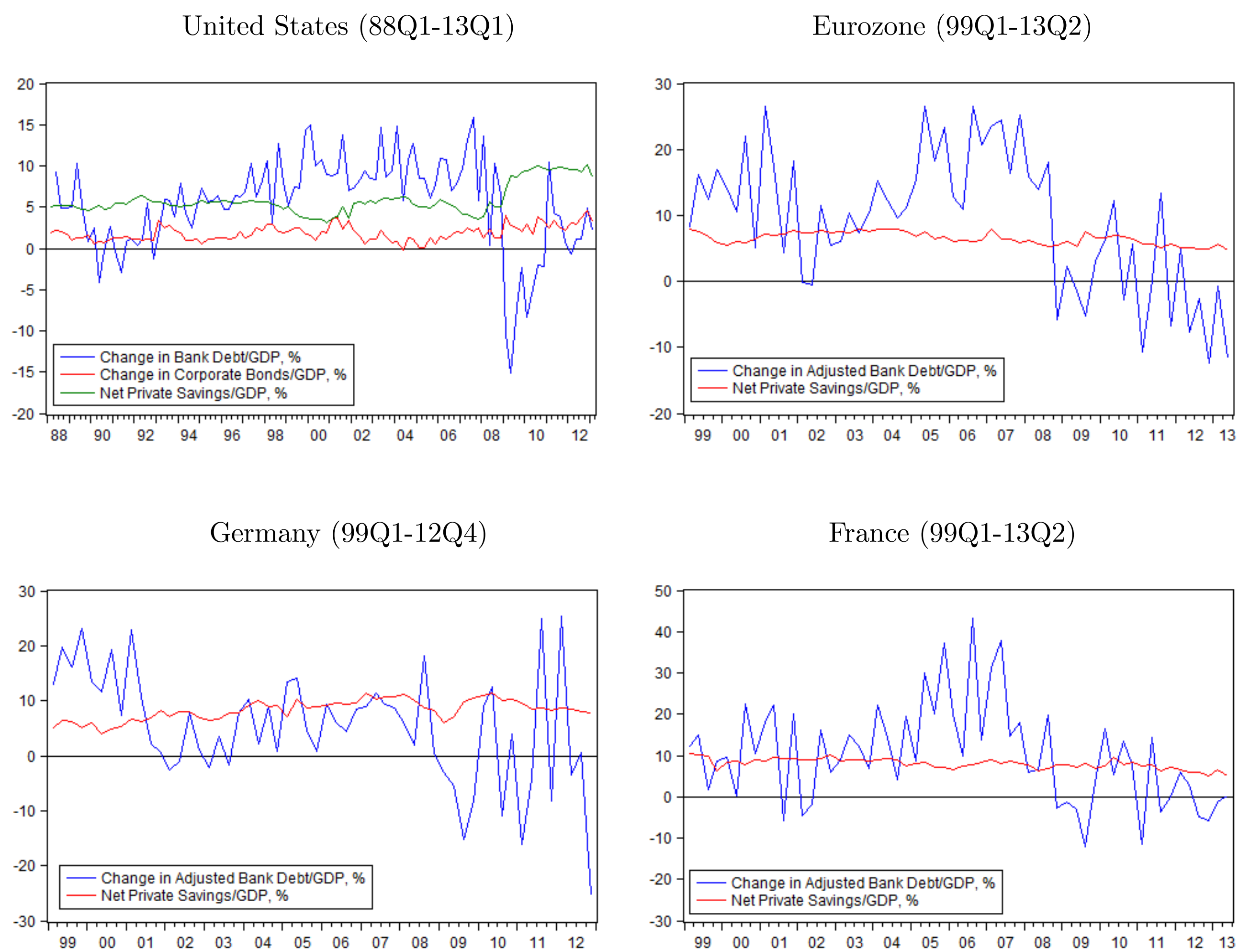
Histogram of 2009Q4 on 2009Q3 elasticity of equity with respect to assets. 200 largest U.S. banks by asset size.

Figure 10. Bank Balance Sheets: Time Series Evidence for Six Countries



Quarter-on-quarter percent changes in aggregate banking system assets, debt and equity.
Data: Flow-of-funds. Each point represents one quarter.
Sample sizes shown in text. p-values of regression slopes in brackets.

Figure 11. Bank Balance Sheets: Changes in Bank Debt versus Net Private Saving



Blue Line: Quarterly changes in aggregate banking system adjusted bank debt, divided by the same quarter's GDP.
Red Line: Quarterly net private saving, divided by the same quarter's GDP.
Green Line: Quarterly change in U.S. corporate bonds, divided by the same quarter's GDP.
Data: Debt from flow of funds (Federal Reserve, ECB), saving from national accounts.

Figure 12. Bank Financing and Bond Financing in the United States

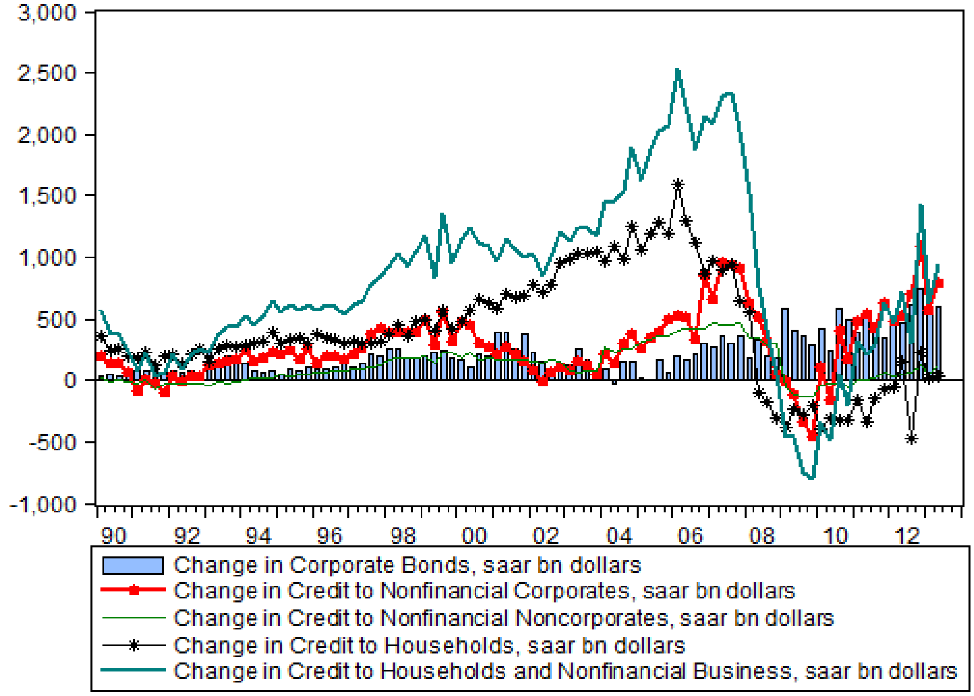


Figure 13. Quantity Rationing, Price Rationing and the U.S. Business Cycle