Money Creation by Banks, Regulation and Optimal long-run Inflation Targets

Christian Wipf

Inauguraldissertation zur Erlangung der Würde eines

Doctor rerum oeconomicarum

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Universität Bern

Bern, Juli 2020

Die Fakultät hat diese Arbeit am 20.8.2020 auf Antrag der beiden Gutachter Prof. Dr. Cyril Monnet und Prof. Dr. Chao Gu als Dissertation angenommen, ohne damit zu den darin ausgesprochenen Auffassungen Stellung nehmen zu wollen.

Selbständigkeitserklärung

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche kenntlich gemacht. Mir ist bekannt, dass andernfalls der Senat, gemäss dem Gesetz über die Universität, zum Entzug des aufgrund dieser Arbeit verliehenen Titels berechtigt ist.

Bern, 20. Juli 2020

Christian Wipf

Acknowledgements

I am extremely thankful to Lukas Altermatt and Mohammed Ait-Lahcen for the exchange, encouragement and friendship with was crucial for me and from whom I learned so much. I also thank Kumar Rishabh for our collaboration and for his assistance in difficult times.

I thank Cyril Monnet for his advice and the lessons about academia and economics, my second-advisor Chao Gu for her feedback and Aleksander Berentsen for inviting me to his workshops in Marrakech and showing me the art of pragmatic but not simplistic feedback and discussion. I also thank Fabrice Collard for his engagement for the PhD-students in Bern and Dirk Niepelt for the discussions and the intellectual and verbal rigour.

I am grateful to my PhD-colleagues at the University of Bern and to my peers and the teachers of my Gerzensee class. I am especially grateful to Tamara Bischof for her kindness and support.

Contents

1 Introduction **1**

2 Should Banks Create Money? **5**

2.1 Environment . . . 8

2.2 Unconstrained fractional reserve banking . . . 11

2.3 Constrained fractional reserve banking . . . 21

2.4 Conclusion . . . 26

3 Liquidity, the Mundell-Tobin Effect, and the Friedman Rule **43**

3.1 The model . . . 48

3.2 Equilibrium with perfectly liquid capital . . . 54

3.3 Equilibrium with perfectly illiquid capital . . . 55

3.4 Equilibrium with partially liquid capital . . . 63

3.5 Conclusion . . . 69

4 Optimal Bank Financing with Less Opaque Assets **83**

4.1 The model . . . 85

4.2 Empirical analysis . . . 99

Introduction

The three chapters of the dissertation cover three distinct topics in monetary macroeconomics and banking. In **chapter 2** I study the welfare implications of money creation by commercial banks and of proposals to limit or prohibit this ability. **chapter 3** is about the optimal long-run inflation targets of central banks and **chapter 4** treats the optimal financing structure of different bank assets like business loans, mortgages or securities. In the following I briefly outline the motivation and the content of the three chapters.

Chapter 2, "Should Banks Create Money?" treats the role of commercial banks in the money supply of an economy.[1] Currently central banks issue cash and reserves and commercial banks issue demand deposits which are claims on central bank money. Typically banks operate under a fractional reserve banking system: they issue demand deposits in excess of the central bank money they hold against redemptions. This system has been consistently criticized, especially after financial crises (1929, 2007/8). Opponents essentially argue that fractional reserve banking (money creation by commercial banks) makes the economy less stable and has no social benefits. As a consequence they want to limit or prohibit this ability of banks with so called "Narrow Banking" proposals. Narrow banks must fully back the money they issue with central bank money and thus the central bank perfectly controls the money supply. An example for a narrow banking proposal is the "Vollgeld" initiative in Switzerland rejected by Swiss voters in June 2018.

In chapter 2 I analyze the welfare implications of such proposals. Abstracting from fragility issues I show that fractional reserve banking (money creation by commercial banks) can be preferable to narrow banking. Under fractional reserve banking the lending of banks is less constrained than in a narrow banking system. More loans increase the return on bank assets and competition forces banks to pass this return to the holders of demand deposits in the form of higher interest payments. In an environment with inflation this is beneficial because inflation acts like a tax on real activity. Fractional reserve banking is beneficial compared to narrow banking because it partially compensates the agents against this "inflation tax" through higher interest payments on demand deposits.

Chapter 3, "Liquidity, the Mundell-Tobin Effect and the Friedman Rule" (co-authored with Lukas Altermatt), was motivated by the mismatch between theory and practice when it comes to op-

[1] Since the late 19th century commercial banks provided the vast majority of the money supply in advanced economies. For example in Switzerland the share of public money (at this time coins issued by the government) sharply decreased since 1850 and private money, banknotes (at this time still issued by commercial banks) and demand deposits increased, see Baltensperger and Kugler [2017].

timal long-run inflation targets. Most monetary macro models find a long-run inflation rate to be optimal where the opportunity costs of holding money are zero (the so called "Friedman rule" after Friedman [1969]) which typically implies a deflationary inflation target. In practice however, no central bank runs the Friedman rule. Actual long-run inflation targets in advanced economies are around 2%. In the paper we want to reconcile theory and practice by providing a theoretical justification for long-run inflation targets above the Friedman rule.

The argument we explore is based on the so called "Mundell-Tobin effect". Mundell [1963] and Tobin [1965] argued that money and capital are to some extent substitutes as a store of value and thus changes in inflation can influence investment. For example if the rate of return of money goes down (inflation goes up) agents substitute away from money into capital, i.e. they invest more. Thus at the Friedman rule, where holding money is costless and we have deflation, agents hold more money and invest less than at higher inflation rates. Could it be that agents hold too much money and too little capital at the Friedman rule and thus a higher inflation rate would increase capital investment and welfare? In a model with a fundamental liauidity-return trade-off between money and capital, i.e. capital has a higher return but money is more liquid, we show that indeed, the optimal long-run inflation rate and the Mundell-Tobin effect are related: When there is a Mundell-Tobin effect an inflation rate above the Friedman rule is optimal and without the Mundell-Tobin effect the Friedman rule is optimal. Thus the Mundell-Tobin effect could be a justification for optimal long-run inflation targets above the Friedman rule.

Chapter 4, "Optimal Bank Financing with Less Opaque Assets" (co-authored with Kumar Rishabh), was motivated by the recent shift in bank loan portfolios from business loans towards mortgages in advanced economies (Jordà et al. [2016]). For example more than 80% of bank loans in Switzerland currently are mortgages. But mortgages and business loans seem to be quite different assets. One essential difference seems to be the ease with which these two assets can be valued by the bank (and by the investors of a bank). The value of a mortgage, which is backed by real estate, mostly depends on publicly observable factors like interest rates or the location of the building and is therefore relatively easy to value. The value of a business loans however, which is mostly backed by the value of a small or medium-sized, unlisted firm, depends more on factors only observable by the firm like the human capital of the entrepreneurs. This makes them more difficult to value for outsiders. This difference is e.g. apparent in the fact that for mortgages there is platform lending and a secondary market while both is not true for business loans. Mortgages are therefore said to be less "opaque" in the language of banking theory.

In chapter 4 we ask the question whether it is useful to finance less opaque assets like mortgages with demandable liabilities (demand deposits) as banks do. In a theoretical model we show the answer is probably no. While demandable liabilities are optimal to finance opaque assets like business loans (in line with the literature on the disciplining role of demandable debt like Calomiris and Kahn [1991]), less opaque assets like mortgages or securities should be financed with non-demandable liabilities like long-term debt or equity. In line with this theory we document a weak positive correlation between opaque assets (business loans) and a suitable measure for demandable liabilities for small and medium sized banks (up to the 75th percentile) using US bank level balance sheet data from 1992 to 2018. But we find no correlation for bigger banks. The reason could be that big banks might enjoy an implicit insurance e.g. in the form of too-big-to-fail guarantees which distorts the choice of their asset and liability structure.

Bibliography

Ernst Baltensperger and Peter Kugler. *Swiss Monetary History since the Early 19th Century.* Studies in Macroeconomic History. Cambridge University Press, 2017.

Charles W Calomiris and Charles M Kahn. The role of demandable debt in structuring optimal banking arrangements. *American Economic Review*, pages 497–513, 1991.

Milton Friedman. The optimum quantity of money. *The Optimum Quantity of Money and Other Essays*, pages 1–50, 1969.

Òscar Jordà, Moritz Schularick, and Alan M. Taylor. The great mortgaging: housing finance, crises and business cycles. *Economic Policy*, 31(85):107–152, 01 2016.

Robert Mundell. Inflation and real interest. *Journal of Political Economy*, 71(3):280–283, 1963.

James Tobin. Money and economic growth. *Econometrica*, 33(4):671–684, 1965.

Should Banks Create Money?

Contemporary monetary systems are characterized by a mixture of public and private means of payment. Central banks issue cash and reserves (outside money) and commercial banks issue demand deposits (inside money). Demand deposits are claims on outside money redeemable on demand. Typically banks operate under a fractional reserve banking system. They issue inside money in excess of the outside money they hold against redemptions. This system has been consistently criticized. Opponents of fractional reserve banking typically argue that it brings instability and question social benefits. They propose to separate the monetary function of banks from their other functions. If banks issue inside money they should back it fully with very safe and liquid assets while other assets should be funded by non-monetary liabilities like long-term debt or equity. The most prominent example of such a "narrow banking" proposal is the "Chicago plan" from 1933 which called for a full backing of inside money by outside money (reserves). A similar plan was advocated by Friedman [1960] and related proposals came up after the recent financial crisis.[1]

The paper aims to analyze the long run welfare implications of such proposals. Although interest in these questions has increased recently, the macroeconomic literature on the topic (where money is explicitly modelled as a nominal asset) is still surprisingly scant given the long history and the fundamental nature of the debate.[2] The specific focus of the paper are the potential benefits of fractional reserve banking in an environment where banks have a "monetary" role, i.e. their liabilities circulate as means of payment (inside money).[3] An emphasis lies on the concrete definition of narrow banking. Narrow banking means that banks must back their demandable (monetary) liabilities fully with very safe and liquid assets but they can acquire other assets if they fund them with non-monetary liabilities. In many models banks are restricted to issue one type of liabilities (an example is Chari and Phelan [2014]) which severely limits narrow banking systems. In the model presented here banks will thus be able to issue other liabilities besides inside money.

I use a basic "New Monetarist" model building on Berentsen et al. [2007]. A preference shock

[1] In Switzerland there was a vote to introduce a narrow banking system in June 2018. See Pennacchi [2012] for an overview of the history of narrow banking and related proposals.

[2] This excludes two- or three-periods real models like Faure and Gersbach [2019], Benigno and Robatto [2019], Jackson and Pennacchi [2019] and Stein [2012]. Wile often providing useful intuitions, these models lack crucial aspects of monetary economies like inflation.

[3] The instability of fractional reserve banking systems has been extensively studied in the literature following Diamond and Dybvig [1983]. However, these models typically ignore the monetary role of bank liabilities and derive their demandable nature from liquidity shocks. A more recent example is Andolfatto et al. [2016]

divides agents into consumers (buyers) and producers (sellers) and buyers acquire money to buy goods from sellers. There is outside money issued by the central bank and inside money issued by perfectly competitive commercial banks. Holding money is costly because the inflation rate lies above the Friedman rule. This "inflation tax" depresses real activity and is the basic inefficiency in the model. The uncertainty from the preference shock aggravates this basic inefficiency. The risk of ending up as a seller with costly idle money balances makes acquiring money even less attractive ex ante.

Banks have two roles in this economy. They issue means of payment (inside money) and they provide liquidity insurance against the preference shock by reallocating money balances after the shock hits. Banks acquire outside money and loans and finance them by issuing inside money and non-monetary debt. Fractional reserve banks partially back their inside money with outside money while narrow banks fully do it. I also consider a "constrained" fractional reserve banking system where banks can only issue inside money and no non-monetary debt.[4]

The analysis shows that fractional reserve banking is beneficial. If banks issue more inside money with respect to outside money welfare rises. This is interesting because the quantity theory of money would predict that the quantity of money and its composition (between inside and outside money) should be irrelevant for the equilibrium real allocation and welfare.

The reason why the quantity theory doesn't apply here are the interest payments on inside money. In equilibrium the interest rate on inside money is a weighted average between the loan rate (which equals the inflation tax by the Fisher equation) and the return on cash (which is one in nominal terms). If banks issue more inside money, lending increases and the asset mix of banks shifts towards more loans. Thus the interest rate on inside money rises. Higher interest payments on inside money induce sellers to produce more for the same price (or to produce the same for a lower price) and this increases the value of real balances for buyers. This is how interest on inside money compensates the agents for the inflation tax and reduces the welfare costs of inflation. The compensation is partial however, because the interest on inside money (the weighted average) is always below the inflation tax.

The mechanism is equivalent to an economy where the central bank pays interest on outside money (see Rocheteau and Nosal [2017]). This equivalence confirms the finding by Brunnermeier and Niepelt [2019] who argue that every fractional reserve banking allocation can be replicated with a narrow banking system when accompanied with appropriate transfers/open-market operations by

[4]This restriction could reflect a distortion in the choice of the liability structure of banks. It could be privately beneficial for banks to choose a higher level of inside money as liabilities than socially optimal (e.g. because of deposit insurance).

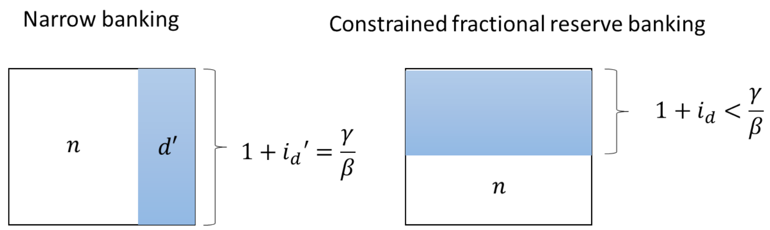
a fiscal authority and a central bank. The result complements the usual argument of models in the spirit of Diamond and Dybvig [1983] where fractional reserve banking is beneficial because it increases high-return long term investment.

The paper also shows that fractional reserve banking dominates narrow banking in terms of welfare. This is not surprising given the first result. Under fractional reserve banking banks could also choose to back the inside money they issue fully with outside money, i.e. they could choose to become narrow banks voluntarily. And under perfect competition private and social interest are typically aligned. Thus the fact that banks do choose to become fractional and not narrow banks indicates that welfare is higher under fractional reserve banking. But narrow banking also improves welfare compared to an economy without banks. Under narrow banking non-monetary debt pays an interest rate equal to the inflation tax. Since this debt is held by sellers after the preference shock they are perfectly compensated for the inflation tax. The same is true under fractional reserve banking, non-monetary debt is also offered at an interest rate equal to the inflation tax. Thus fractional reserve banking achieves the same insurance against the preference shock and on top of that also offers a partial compensation for the inflation tax on inside money by offering a positive interest rate. Narrow banking does not offer this kind of compensation because the interest rate on inside money is zero. The allocation in the narrow banking economy is the same as in the basic model of Berentsen et al. [2007]. Thus the paper shows how the Berentsen et al. [2007] model can be interpreted as a narrow banking economy.

There is a debate on whether it matters if banks are modelled as institutions who influence the money supply by issuing partially backed inside money“ex nihilo” or as institutions who only intermediate already existing funds like outside money. This is exactly the difference between fractional reserve and narrow banks in the model. Fractional reserve banks create inside money “ex nihilo” and influence the money supply while narrow banks can be seen as pure intermediators of outside money as in Berentsen et al. [2007]. Thus the paper shows that it can make a difference how exactly banks are modelled and provides a counterexample to Andolfatto [2018] who finds no substantial effect in his model.

Finally the paper shows that if fractional reserve banking is constrained i.e. if fractional reserve banks can only issue inside money but no non-monetary debt, fractional reserve banking only dominates narrow banking if banks can issue a sufficiently high quantity of inside money. In this economy fractional reserve banks only issue inside money at an interest rate below the inflation tax. In this case sellers holding inside money after the preference shock are only imperfectly compensated for the inflation tax. Thus the insurance against the preference shock is imperfect in contrast to the narrow banking economy where sellers can still use non-monetary debt at an interest rate equal to the inflation tax. On the other hand fractional reserve banking still has the advantage of providing a partial compensation against the inflation tax which narrow banking has not since interest on inside money is zero.

The following figure highlights this difference. In the narrow banking economy we have a full compensation against the inflation tax on the cash deposited by agents against non-monetary debt d' but no compensation on inside money n. In the constrained fractional reserve banking economy we have a *partial* compensation of the inflation tax (because $1 + i_d$ is below the inflation tax) on the *full* stock of inside money n. If fractional reserve banks are sufficiently constrained in issuing inside money the interest rate on inside money is low and narrow banks can yield higher welfare.



The rest of the paper is organized as follows: Section 2.1 shows the basic environment. Then the the full model (section 2.2) and the model with constrained fractional reserve banking (section 2.3) are presented.

2.1 Environment

Basic structure: The environment follows a standard model in the style of Lagos and Wright [2005] as presented in Berentsen et al. [2007]. Time is discrete and continues forever. Every period is divided into two sequential competitive markets called *first* and *second* market. There is a perishable consumption good produced and consumed in both markets denoted q in the first market and x in the second.

Agents: There is a unit mass of infinitely lived agents. They discount future periods with β and they cannot commit. At the beginning of every period agents face a *preference shock* which determines what they can do in the first market. With probability $s \in (0, 1)$ an agent is a *seller* and can only produce and with the inverse probability $1 - s$ an agent is a *buyer* and can only consume. Sellers have a (weakly) convex disutility of production $c(q)$ and buyers utility of consumption is strictly concave $u(q)$ and satisfies the Inada-conditions. In the second market all agents can consume and produce and their preferences are represented by a utility function $x - h$, i.e. they consume(produce) with linear (dis)utility. Denote the efficient quantity of buyer consumption in the first market with q^* given by:

$$\frac{u'(q^*)}{c'(\frac{1-s}{s}q^*)} = 1 \tag{2.1}$$

where $q^* = \frac{s}{1-s}q_s^*$.

Outside money, monetary policy and prices: There is a stock M of outside fiat money called "cash" issued by the central bank evolving at rate $\gamma > 0$, i.e. $M = \gamma M_{-1}$. The growth rate of the cash supply γ is the monetary policy tool of the central bank. She manages the cash supply by lump-sum cash transfers τ to agents in the second market. Since agents have unit mass the transfer/tax per agent is $\tau = M - M_{-1} = (\gamma - 1)M_{-1}$.

Let p be the price of consumption good q in terms of money in the first market and ϕ is the value of fiat money in the second market in terms of consumption goods x (i.e. the inverse of the price level in the second market). Denote (gross) inflation π as the ratio of the prices between two consecutive second markets, i.e. $\pi = \frac{\phi}{\phi_{+1}}$.

Banks, financial contracts and inside money: There is also an infinite amount of perfectly competitive, profit-maximizing firms (banks). In contrast to agents, they can commit and monitor other agents at no cost. The first property enables them to issue debt and the second property enables them to make loans.

Banks can issue two types of debt: *inside money* and *non-monetary debt.* Inside money is debt usable as means of payment in the first market. Banks need to back it with a fraction $\alpha \in (0, 1)$ in outside money. This constraint should capture the idea that the transactions with inside money in the first market generate redemptions from in- and outflows of inside money between banks and to settle and clear these flows banks need some outside money[5] Non-monetary debt is not usable as means of payment in the first market (suppose e.g. it has a longer maturity). But banks don't need to back it with outside money. Both types of debt are nominal, interest bearing claims on outside money. Also both types of debt are fully redeemed in the following second market after they are issued. Bank loans are inside money loans, which are also paid back in the next second market, denominated in outside money.[6] Bank contracts are formed in a *banking period* after the preference shock.

[5]In appendix A.4 I present a model where this constraint arises from redemptions from inside to outside money before the first market.

[6]With linear utility in the second market there is no gain from spreading the redemption of debt or the repayment of loans over multiple periods. Thus assuming this kind of contracts is not constraining in this environment.

Role for money and banks: The role for money in this environment is motivated by limited commitment and anonymity. Since agents cannot commit and are anonymous, buyers cannot issue debt in the first market and sellers require immediate compensation for the goods they produce. Buyers must give sellers "something" if they want to consume in the first market. This why agents (buyers) hold (inside or outside) money.

Banks have two roles in this environment. They can issue liabilities that circulate as means of payment in the first market, i.e. they can create money. And they can reallocate money after the preference shock which is valuable because acquiring money is costly, i.e. banks provide insurance against liquidity risk.

Equilibrium: I focus on stationary and symmetric equilibria. In a stationary environment the value of aggregate real balances is constant over time implying $\phi M = \phi_{+1} M_{+1}$. Since the stock of cash grows at γ also the price level in the second market must grow at γ or $\phi/\phi_{+1} = \gamma$ in a stationary equilibrium. By setting γ the central bank can thus also determine long-run inflation and γ can also be interpreted as the long run inflation target of the central bank.

Throughout the paper I assume holding money is costly, i.e. $\gamma > \beta$ and the central bank does not follow the Friedman rule. The assumption introduces a basic inefficiency into the environment in the form of an "inflation tax" which banks can potentially alleviate. Agents hold too little money for first best consumption in the first market, i.e. inflation acts like a tax on consumption/production in the first market. Since inside money is a claim on outside money the inflation tax also applies to inside money. The preference shock aggravates this basic inefficiency from the inflation tax. If acquiring money is costly, the risk to be a seller with (costly) idle money holdings in the first market makes acquiring money even less attractive ex-ante.

Sequence of events: Figure 2.1 summarizes the sequence of events in this economy: Agents acquire outside money in the second market. In the banking period after the preference shock, they can can deposit this outside money in banks and borrow to acquire inside money and non-monetary debt. In the first market they consume and produce and finally in the second market all inside money and all non-monetary debt is redeemed and the loans are repaid.

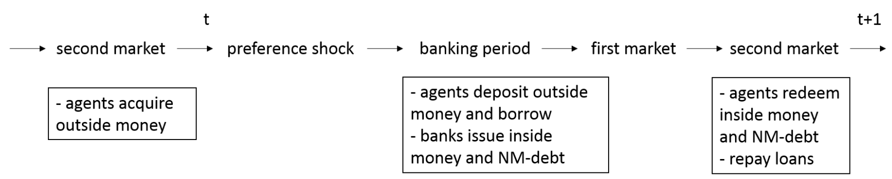


Figure 2.1: sequence of events

2.2 Unconstrained fractional reserve banking

2.2.1 Banks

Banks issue inside money and non-monetary debt in the banking period. They issue inside money against cash deposits d and as loans l. They must back this inside money with at least a fraction $\alpha \in (0, 1)$ of outside money. The interest on inside money is i_d and the interest on loans is i. Banks can also issue non-monetary debt against cash deposits d' at interest i'_d. A representative bank maximizes the nominal value of her assets (cash and loans) minus the value of liabilities (inside money, non-monetary debt) subject to having enough cash to satisfy the reserve constraint. The problem of a representative bank is:

$$\max_{l,d,d'} = d + d' + l(1+i) - (l+d)(1+i_d) - d'(1+i'_d) \tag{2.2}$$

$$s.t. \quad \alpha(l+d) \leq d + d'$$

If $i > i_d$ and $i_d, i'_d > 0$ the bank would like to make a loan as big as possible and to have as little cash deposits as possible. The reserve constraint will bind and the bank will not hold excess-reserves. Suppose this holds, then we get the following relationship between loans and cash deposits:

$$l = \frac{d'}{\alpha} + \frac{1-\alpha}{\alpha} d \tag{2.3}$$

Note that the loan size increases exponentially as α decreases for a given d or d' and the increase is stronger for a marginal increase in d' than in d. For example if $\alpha = 0.5$ and the bank takes a cash deposit against inside money, her lending capacity increases by 1. If the bank instead takes a cash deposit against non-monetary debt, her lending capacity increases by two. This is because issuing inside money triggers further cash acquisitions over the reserve constraint who converge to $\frac{\alpha}{1-\alpha}$ in the end while issuing non-monetary debt does not have this consequence.

Using the binding reserve constraint we can rewrite the objective function as:

$$\max_{d,d'} \quad \frac{d}{\alpha}(\alpha + (1-\alpha)(1+i) - (1+i_d)) + \frac{d'}{\alpha}(\alpha + (1+i) - \alpha(1+i'_d) - (1+i_d))$$

Thus if banks use both types of debt and we apply free entry (zero profits) we find that the interest rate on inside money and the interest rate on non-monetary debt satisfy:

$$1 + i_d = \alpha + (1 - \alpha)(1 + i) \tag{2.4}$$

$$1 + i'_d = 1 + i \tag{2.5}$$

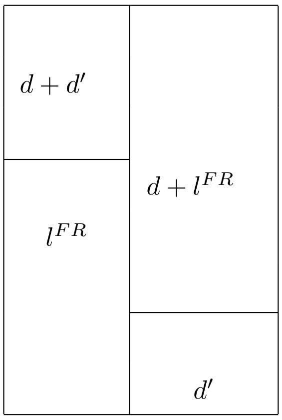
Thus the interest rate on inside money i_d is a weighted average of the return on cash (1 in nominal terms) and the return on loans $(1 + i)$. As the gross loan rate will be bigger than one, the interest on inside money must be below the loan rate. This also implies it is below the interest on non-monetary debt, $i'_d > i_d$.

Also i_d increases if α goes down. This is because a lower α shifts the asset mix of the bank from assets with no return (cash) to assets with return (loans). Consequently, the bank pays an interest on its liabilities (inside money) closer to the loan rate under zero profits. The spread between loan rate and interest on inside money decreases. For example if α decreases from 0.5 to 0.25 and the loan rate stays constant at $i = 0.2$ the return on inside money increases from $i_d = 0.1$ to 0.15. If $\alpha \to 1$ the interest on inside money goes to zero. This is the case of narrow banking where issued inside money must be backed fully with outside money. If $\alpha \to 0$ the interest on inside money approaches the loan rate. In this case banks don't need to back inside money with outside money thus for the bank there is no difference between issuing inside money and non-monetary debt. i'_d does not depend directly on α.

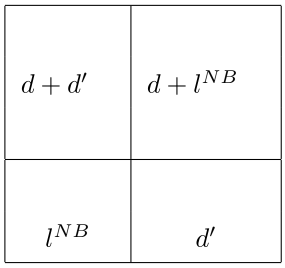
The two types of debt yield the following trade-off for the bank: Non-monetary debt has the advantage that it increases cash holdings without increasing inside money (which would trigger further cash holdings). Thus loans can increase by $1/\alpha$ at the margin with non-monetary debt while in the case of deposits against inside money they can only increase by $(1 - \alpha)/\alpha$. The disadvantage of non-monetary debt are the higher funding costs since $i'_d > i_d$.

We define a *fractional reserve banking system* as an economy where banks don't fully back their issued inside money with outside money, i.e. $\alpha \in (0, 1)$. A *narrow banking system* is an economy where bank fully back their issued inside money with outside money, i.e. $\alpha = 1$. If banks acquire loans they must fund them with non-monetary debt. The following figure shows the balance sheets of a fractional reserve and a narrow bank for the same amount of deposits against inside money d and non-monetary debt d'.

fractional reserves



narrow bank



In a fractional reserve banking system the issued inside money $(d+l^{FR})$ exceeds the outside money deposited $(d+d')$ while under narrow banking they must be equal by definition. This implies that fractional reserve banks can lend more $(l^{FR} > l^{NB})$ because they don't have to back inside money $1:1$ with outside money. Narrow banks' lending capacity is constrained by the cash deposits against non-monetary debt $l^{FR} = d'$. But fractional reserve banks can lend according to (A.15). The last difference concerns the interest rate on inside money. Under narrow banking banks only accept cash deposits for inside money d if they pay zero interest, i.e. if $i_d = 0$, see (2.4). Otherwise they would set $d = 0$. Thus in a narrow banking system inside money and outside money are perfect substitutes.

2.2.2 Second market

A representative agent may bring outside money (m) inside money n, non-monetary debt d' and some own debt l into the second market. He chooses consumption x, work h and his new holdings of outside money m_{+1}. $V(m_{+1})$ denotes the expected value of entering the next period with m_{+1} units of outside money where $V(m_{+1}) = sV_s(m_{+1} + (1-s)V_b(m_{+1})$ i.e. the expected value of entering next period with m_{+1} units of money is the value as a buyer/seller times the respective probabilities.

$$W(m, n, d', l) = \max_{x,h,m_{+1}} x - h + \beta V(m_{+1}) \tag{2.6}$$

$$s.t. \quad x + \phi m_{+1} = h + \phi(\tau + m) + n\phi(1 + i_d) + d'\phi(1 + i'_d) - l\phi(1 + i)$$

The first order condition for optimal (positive) outside money holdings solve:

$$\phi = \beta V'(m_{+1}) = sV_s'(m_{+1} + (1-s)V_b'(m_{+1}) \tag{2.7}$$

(2.7) implies that agents want to choose the same amount of cash to bring into the next period - independent of m, n, d' and l. This is a consequence of the linear utility function introduced by Lagos and Wright [2005]. The envelope conditions to the problem are:

$$\begin{aligned} W_m &= \phi \\ W_n &= \phi(1 + i_d) \\ W_d' &= \phi(1 + i_d') \\ W_l &= -\phi(1 + i) \end{aligned} \tag{2.8}$$

Finally market clearing for the output good and for money solves:

$$(1-s)h_b + sh_s = (1-s)x_b + sx_s \tag{2.9}$$

$$m_{+1} = M \tag{2.10}$$

2.2.3 Banking period and first market

I focus on the case where agents use only inside money in the first market. In appendix A.3 I show that buyers will strictly prefer to acquire inside money instead of outside money if interest on inside money is positive, i.e. if $i_d > 0$ and I assume they also acquire inside money if $i_d = 0$. Since after the preference shock all uncertainty is resolved, the problem of the banking period and the first market for buyers or sellers can be taken together.

buyer problem

A buyer arrives with m units of outside money in the banking period. There he decides how much of this he should deposit for inside money d_b and for non-monetary debt d_b' and how much he should borrow l_b. Then, in the first market he chooses how much to consume q_b given the amount of inside money $n = d_b + l_b$ he has.

$$V_b(m) = \max_{q_b, l_b, d_b, d_b'} \quad u(q_b) + W(m - d_b - d_b', l_b + d_b - pq_b, d_b', l_b) \tag{2.11}$$

$$s.t. \quad pq_b \leq d_b + l_b$$

$$d_b + d_b' \leq m$$

It is clear that the buyer should deposit all his cash be it for inside money or non-monetary debt, i.e. the second constraint must bind and $d_b' = m - d_b$. From the envelope conditions (2.8) the marginal value of inside money and non-monetary debt dominate the marginal value of cash. Second, I focus on an interior solution for borrowing. A buyer would not borrow if he already brings sufficient outside money balances m for his unconstrained level of consumption. But this could only happen if acquiring money is costless, i.e. if the inflation tax is zero or $\gamma = \beta$ which is not what we assume. The problem yields the following first-order conditions (also using (2.8) and with λ denoting the multiplier for the constraint):

$$\begin{aligned} q_b: &\quad u'(q_b) = p(\phi(1+i_d) + \lambda) \\ l_b: &\quad \lambda + \phi(1+i_d) = \phi(1+i) \\ d_b: &\quad \lambda + \phi(1+i_d) \geq \phi(1+i_d') \end{aligned}$$

The last constraint is formulated with a weak inequality meaning that if the inequality is strict the buyer wants to choose $d_b = m$ and $d_b' = 0$. From the banking problem we know that $i > i_d$ in equilibrium. Thus the constraint in the first market must bind and the buyer will be liquidity constrained, $\lambda = \phi(i - i_d)$. We also know from the banking problem that $i = i_d'$ in equilibrium. This implies the buyer is indifferent between depositing his outside money for inside money or for non-monetary debt and any combination of $d_b + d_b' = m$ is fine. The third condition holds at equality. Without loss of generality we will assume that the buyer deposits all his cash for non-monetary debt. Thus the solution to problem (2.11) is given by:

$$u'(q_b) = p\phi(1+i) \tag{2.12}$$

$$l_b = pq_b \tag{2.13}$$

$$d_b = 0 \quad , \quad d_b' = m \tag{2.14}$$

And the marginal value of outside money for a buyer is:

$$V_b'(m) = \phi(1+i_d') \tag{2.15}$$

seller problem

A seller also arrives with m units of outside money in the banking period. He can deposit his outside money for inside money d_s or for non-monetary debt d_s' and he can borrow l_s. In the first market he chooses production q_s.

$$V_s(m) = \max_{q_s,l_s,d_s,d'_s} \quad -c(q_s) + W(m - d_s - d'_s, d_s + l_s + pq_s, d'_s, l_s) \tag{2.16}$$
$$s.t. \quad d_s + d'_s \leq m$$

The envelope conditions (2.8) and the relations on interest rates derived in the banking problem $i = i'_d > i_d$ significantly simplify the analysis. Also sellers will deposit all their cash for inside money or non-monetary debt. But since inside money has no liquidity value for they they strictly prefer to deposit for non-monetary debt, i.e. $d'_s = m$ and $d_b = 0$. Also sellers don't borrow if $i > i_d$, so $l_s = 0$. Thus the optimality conditions for the sellers are

$$c'(q_s) = p\phi(1 + i_d) \tag{2.17}$$
$$l_s = 0 \tag{2.18}$$
$$d_s = 0 \quad , \quad d'_s = m \tag{2.19}$$

and the marginal value of outside money for a seller is:

$$V'_s(m) = \phi(1 + i'_d) \tag{2.20}$$

Finally we have the market clearing conditions in the first market. Denote total bank demand for deposits against non monetary debt as d' and total bank demand for deposits against inside money as d and total bank supply of loans as l. Using the optimality conditions from above we have the following market clearing conditions in the first market:

$$d' = (1 - s)d'_b + sd'_s = m$$
$$d = (1 - s)d_b + sd_s = 0$$
$$l = (1 - s)l_b$$
$$(1 - s)q_b = sq_s \tag{2.21}$$

Combine this with the binding reserve constraint of banks to get:

$$(1 - s)l_b = \frac{m}{\alpha} \tag{2.22}$$

2.2.4 Equilibrium

We first solve for the equilibrium interest rates. We combine the expressions for the marginal value of outside money for a buyer (2.15) and for a seller (2.20) with the condition for optimal outside money holdings (2.7) to get:

$$\phi = \beta\phi_{+1}(1 + i_{d'+1}) \tag{2.23}$$

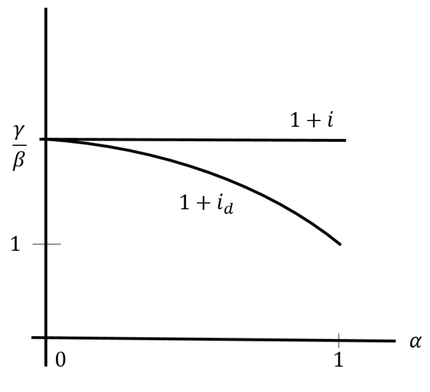
To get the equilibrium interest rates we apply stationarity ($\gamma = \phi/\phi_{+1}$) to (2.23) and use the relations on interest rates from the bank problem, (2.4) and (2.5).

$$1 + i = \frac{\gamma}{\beta} \tag{2.24}$$

$$1 + i'_d = \frac{\gamma}{\beta} \tag{2.25}$$

$$1 + i_d = (1 - \alpha)\frac{\gamma}{\beta} + \alpha \tag{2.26}$$

Thus the loan rate and the interest on non-monetary debt must equal the inflation tax and are independent of α. This can be interpreted as a Fisher equation: the interest rates are the real interest rate ($1/\beta$) times inflation ($\phi/\phi_{+1} = \gamma$). The interest on inside money is then the weighted average of the inflation tax and the return on cash and is thus below the inflation tax and decreasing in α. We get the following picture for the evolution of the interest rates as a function of α:



To get equilibrium consumption in the first market combine optimal consumption (2.12) with optimal production (2.17) and use the equilibrium expressions for the interest rates and market clearing (2.21):

$$\frac{u'(q_b)}{c'(\frac{1-s}{s}q_b)} = \frac{1+i}{1+i_d} = \frac{\gamma/\beta}{(1-\alpha)\gamma/\beta+\alpha} \tag{2.27}$$

The following proposition summarizes the most important results:

Proposition 2.1. *Suppose holding money is costly (i.e. $\gamma > \beta$) and the reserve constraint is interior (i.e. $\alpha \in (0,1)$), then there is a unique stationary equilibrium of an economy with banks, inside money and non-monetary debt in which:*

i) first market consumption solves (2.27)

ii) first market consumption is below first best q^* and the inefficiency increases in the inflation tax γ/β and the reserve constraint α but is independent of the preference shock s.

iii) welfare is higher than in an economy without banks (see A.1), i.e. fractional reserve banking and inside money creation are essential.

iv) as $\alpha \to 1$ (the economy becomes a narrow banking economy) the allocation approaches an economy without banks and preference shock as in A.2.

v) as $\alpha \to 0$ the allocation approaches the first best allocation (2.1).

First the proposition shows that inside money is not neutral in this economy (*ii*)). The more inside money relative to cash (the lower α) the higher consumption in the first market and the higher welfare. Thus the quantity theory of money claiming that the quantity and the composition of inside and outside money are irrelevant does not hold here. This implies that fractional reserve banks are essential, i.e. they improve the allocation compared to an economy without banks and inside money creation (*iii*)).

Why is more inside money (a lower α) beneficial? In section 2.2.1 on banks we saw how a lower α leads to increases in lending and a shift in the asset-mix of banks from assets with no return (cash) to return bearing assets (loans) which allows banks to pay higher interest on inside money. This is beneficial because this interest (partially) compensates the agents for the inflation tax, which is the basic inefficiency in the economy. Since the equilibrium interest rate on inside money is below the inflation tax in equilibrium ($1+i_d < \gamma/\beta$) this compensation is always partial. The mechanism is identical to an economy with only outside money where the central bank pays interest on cash in the second market.[7]

[7] The equilibrium allocation of an economy with interest on outside money is given by $\frac{u'(q)}{c'(q)} = \frac{\gamma}{\beta}\frac{1}{1+i_m}$ where $1+i_m$ is the gross interest on cash by the central bank, see Rocheteau and Nosal [2017] p.140. Note that this

Second: since welfare decreases in α fractional reserve banking with $\alpha \in (0,1)$ dominates narrow banking with $\alpha = 1$ in terms of welfare. This makes intuitive sense. Under fractional reserve banking banks could always decide to become narrow banks voluntarily, i.e. choose to hold more outside money than they are obliged to. Since perfect competition aligns private and social interests the fact that banks don't do this under fractional reserve banking indicates that fractional reserve banking is welfare improving.

However, also narrow banking is essential, i.e. welfare under narrow banking is higher than without banks as in appendix A.1. We saw in the section on banks that narrow banking cannot offer interest on inside money, $i_d = 0$. So narrow banks cannot provide a compensation against the inflation tax and the welfare costs of inflation. But they are useful because they perfectly insure agents against the preference shock. The easiest way to see this is by verifying that the allocation of the narrow banking model is exactly the same as in an economy without preference shock and banks ((iii)), see A.1. The intuition is that since the deposit rate on non-monetary debt exactly equals the inflation tax ($1 + i'_d = 1 + i = \frac{\gamma}{\beta}$) agents that turn out to be sellers are perfectly compensated for the inflation tax on the cash they acquired. So the risk of being a seller who cannot use the cash in the first market disappears and the allocation-worsening role of the preference shock is eliminated. Narrow banks can be seen as a substitute for a market to borrow/lend cash after the preference shock.[8]

The narrow banking is equivalent to an allocation with only outside money and banks reallocating this outside money after the preference shock. This is the basic version of Berentsen et al. [2007]. Thus the proposition shows how the model of Berentsen et al. [2007] can be interpreted as a narrow banking economy where banks issue fully backed inside money and non-monetary debt.[9]

Finally a few comments on the equilibrium if $\alpha = 0$ (v). In this case banks don't need outside money to back the inside money they issue. This means they are only willing to take cash deposits if the interest rate on inside money (and non-monetary debt) is zero, i.e. $i_d = i'_d = 0$. Zero profits then implies that also the loan rate is zero. However, the condition on agents optimal outside money holdings, (2.23), tells us that if interest rates are zero agents would not be willing to hold outside money if the inflation tax is positive ($\gamma > \beta$). Thus this cannot be an equilibrium. Now suppose equilibrium interest rates are $i = i_d > 0$ and $i'_d > 0$ At these interest rates bank demand for cash deposits (either for inside money or non-monetary debt) is zero. They only make loans in inside money and since $i = i_d$ they also make zero profits. However, for this equilibrium to exist

expression is exactly identical to equation 2.27.

[8]See [Rocheteau and Nosal, 2017], chapter 8.5 for this equivalence.

[9]In footnote 9 of Berentsen et al. [2007] the authors also make the interpretation of their model as a narrow banking economy.

also the supply of cash deposits must be zero i.e. agents don't want to hold cash anymore. This is satisfied if the marginal costs of holding outside money are higher than the marginal benefits, i.e. if (2.23) is an inequality. In equilibrium we must thus have that both interest rates on inside money and on non-monetary debt are positive but below the inflation tax, i.e.

$$0 < i_d, i_d' < \gamma/\beta - 1 \tag{2.28}$$

Any interest rates satisfying (2.28) would be an equilibrium if $\alpha = 0$. In such an economy outside money has no function anymore, it is a pure inside money economy.[10] Because the loan rate and the interest on inside money are identical buyers are never liquidity constrained and their holdings of inside money can be anything from the quantity to consume the first best to infinity, i.e. $l_b \in (pq^*, \infty)$. The economy achieves the first best allocation equivalent to an economy with direct credit.

2.2.5 The non-neutrality of inside money

To see the beneficial effects of higher interest on inside money more clearly look at supply and demand for the consumption good in the first market. Equilibrium supply from the seller side is implicitly defined by optimal production (2.17), $c'(q) = p\phi(1 + i_d)$, and equilibrium demand is given by the buyer optimality condition $u'(q) = \gamma/\beta p\phi = (1 + i)p\phi$. Figure 2.2 shows demand and supply according to these equations. Specifically it depicts a shift from initial supply $S(\alpha_1)$ to the new supply curve $S(\alpha_2)$ for a decrease in α from an arbitrary value α_1 to a lower value α_2. Note that supply weakly increases in the interest on inside money from the properties of the utility functions. Higher interest payments on inside money induce the sellers to produce more for the same relative price (or to produce the same for a lower price). The figure shows that as a result of this change equilibrium consumption (and production) will clearly increase and the relative price $p\phi$ decreases.

[10] Since all prices and contracts were defined in outside money such an economy would have to use another numeraire.

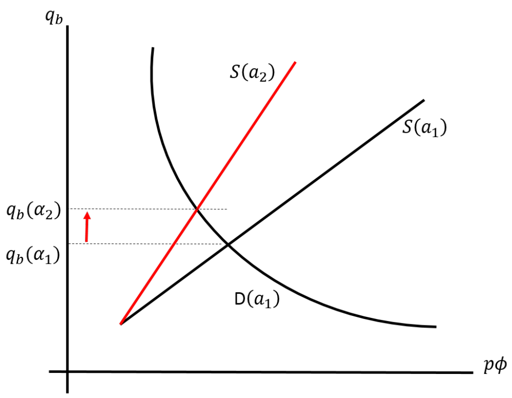


Figure 2.2: The equilibrium effects of an increase in interest of inside money

How can we see the non-neutrality from a quantity theory perspective? Look at the inside money a buyer brings to the first market. In equilibrium a buyer deposits his outside money for non-monetary debt and acquires inside money only by borrowing. Thus his inside money holdings are just l_b. Using market clearing in the first and the second market, (2.22) and (2.9), we get that the inside money holdings of a buyer are $l_b = \frac{M_{-1}}{\alpha(1-s)}$. Equilibrium consumption in the first market can then be written as:

$$q_b = \frac{n}{p} = \frac{M_{-1}/(\alpha(1-s))}{p} \tag{2.29}$$

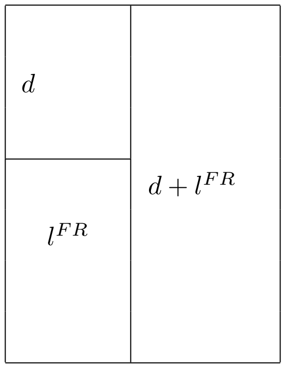
If the quantity theory would hold with respect to inside money this would imply that the price level in the first market rises 1 : 1 with the amount of inside money available for the buyer. For example if α decreases from 0.5 to 0.25 the amount of inside money available for a buyer l_b more than doubles. However, since also interest on inside money goes up (per unit) sellers accept this amount of inside money at a lower price than a 1 : 1 increase. Thus the denominator of (2.29) rises less than the denominator of (2.29) and real consumption q_b is not constant but rises.

2.3 Constrained fractional reserve banking

In this section, I assume banks are restricted to issue inside money under fractional reserve banking. But the narrow banking allocation is the same as before. This shows how the fact that banks can issue different types of liabilities matters for the results. The allocation could be interpreted as a situation where the private interests of banks and the social interests are not aligned and banks - for some reason - have a private benefit of issuing more (or only) inside money (e.g. because it's cheaper) than socially optimal i.e. the choice of the liability structure of banks is distorted. The balance sheet of banks with only inside money as liabilities is:

constrained

fractional reserves



For banks this means d' is zero in (2.2) and we only have i_d and i as interest rates. If $i > i_d > 0$ also here the bank wants to make a loan as big as possible and to hold as little outside money as possible. Thus the reserve constraint will bind and instead of (A.15) we get

$$l = \frac{1-\alpha}{\alpha}. \tag{2.30}$$

Then, under zero profits (2.4) still holds.

$$1 + i_d = \alpha + (1-\alpha)(1+i) \tag{2.4}$$

For buyers and sellers the basic problem is still (2.11) and (2.16). For them the difference is that they cannot deposits their outside money for non-monetary debt anymore. In this case both will deposit their outside money for inside money and we have:

$$d_b = d_s = m \tag{2.31}$$

The change also affects the marginal value of outside money. For buyers (2.15) becomes now:

$$V_b'(m) = \frac{u'(\frac{m+l}{p})}{p} = \phi(1+i) \tag{2.32}$$

and for sellers (2.20) becomes

$$V_s'(m) = \phi(1 + i_d) \tag{2.33}$$

since they now deposit for inside money and not for non-monetary debt. This has implications for optimal outside money holdings of buyers. Instead of (2.23) (2.7) now yields:

$$\phi = \beta\phi_{+1}[(1 - s)(1 + i_{+1}) + s(1 + i_{d+1})]. \tag{2.34}$$

or under stationarity:

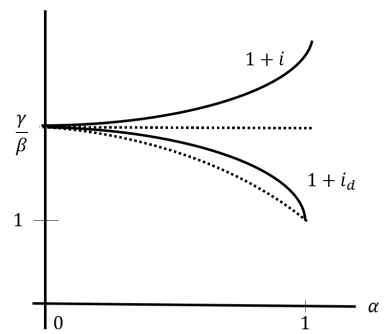
$$\frac{\gamma}{\beta} = (1 - s)(1 + i) + s(1 + i_d) \tag{2.35}$$

Combining (2.35) with the relation for interest rates from the bank problem (2.4) yields the following equilibrium interest rates:

$$1 + i_d = \frac{(1 - \alpha)\gamma/\beta + \alpha(1 - s)}{1 - \alpha s} \tag{2.36}$$

$$1 + i = \frac{\gamma/\beta - \alpha s}{1 - \alpha s} \tag{2.37}$$

The big difference to before is that now the loan rate increases with α (before it was independent of α). This is because before both agents could deposit their outside money at rate i_d' which was equal to i. Thus it was irrelevant whether an agent turned out to be a buyer or a seller and $1 + i_d' = 1 + i = \gamma/\beta$ in equilibrium. Now since there is a spread between the loan rate and the deposit rate $i > i_d$ the preference shock matters because as a seller an agent gets less. Thus as a buyer the agent must be compensated with an interest $1+i > \gamma/\beta$ in equilibrium and $1+i_d < \gamma/\beta$. As i_d is decreasing in α this also implies $1 + i$ must rise with α otherwise the agent would not hold outside money in equilibrium. The following figure shows this evolution in equilibrium (the dashed lines are the interest rates of the economy without preference shock).



To get equilibrium consumption in the first market we again combine the expressions for equilibrium interest rates with optimal buyer consumption (2.12) and seller production (2.17).

$$\frac{u'(q_b)}{c'(\frac{1-s}{s}q_b)} = \frac{1+i}{1+i_d} = \frac{\gamma/\beta - \alpha s}{(1-\alpha)\gamma/\beta + \alpha(1-s)} \tag{2.38}$$

The following proposition summarizes the most important results:

Proposition 2.2. *Suppose holding money is costly (i.e. $\gamma > \beta$) and the reserve constraint is interior (i.e. $\alpha \in (0,1)$), then there is a unique stationary equilibrium with only inside money in which:*

i) first market consumption solves (2.38)

ii) first market consumption is below first best consumption q^* and the inefficiency increases in the inflation tax γ/β, the reserve constraint α and the fraction of sellers in the economy s.

iii) welfare is lower than in the economy with both types of debt from proposition 4.1.

iv) welfare is higher than in an economy without banks as in A.1 and approaches this allocation as $\alpha \to 1$.

v) welfare is higher than in the narrow banking economy from proposition 4.1 if $\alpha < \tilde{\alpha}$ and lower if $\alpha > \tilde{\alpha}$ where $\tilde{\alpha} = \frac{\gamma/\beta}{\gamma/\beta + s} > 0.5$.

vi) as $\alpha \to 0$ the allocation approaches the first best allocation.

The most important result from this proposition is that constrained fractional reserve banking doesn't strictly dominate narrow banking in terms of welfare as before. From *iv*) the allocation with constrained fractional reserve banking approaches an economy without banks and we still know from proposition 4.1 that welfare with narrow banking is higher than in an economy without banks. As welfare increases when α decreases and approaches the first best allocation we have a threshold result. If banks can issue a sufficiently high quantity of inside money ($\alpha < \tilde{\alpha}$) welfare under constrained fractional reserve banking is higher while if banks are very constrained in the issuance of inside money ($\alpha > \tilde{\alpha}$) welfare under (unconstrained) narrow banking is higher.[11]

Remember that under narrow banking inside money pays zero interest, $i_d = 0$ but non-monetary debt pays an interest equal to the inflation tax $1 + i'_d = \gamma/\beta$. Thus agents who turn out to be sellers can deposit their outside money after the preference shock at an interest rate which perfectly

[11] $\tilde{\alpha}$ must be above 0.5 since this is the number reached as $\gamma/\beta \to 1$ and $s \to 1$. So for any $\alpha < 0.5$ fractional reserves is always better.

compensates them for the inflation tax. Thus narrow banking achieves perfect insurance against the preference shock (the basic inefficiency of the inflation tax on inside money however, is not addressed since inside money pays no interest). Note that now with constrained fractional reserve banking this is not the case. Agents who turn out to be sellers now deposit their outside money at $1 + i_d < \gamma/\beta$ since banks can only issue inside money. Thus the insurance against the preference shock is imperfect under constrained fractional reserve banking and narrow banking has a relative advantage in this respect. This explains why the threshold $\tilde{\alpha}$ decreases in s. The higher the risk of becoming a seller the more valuable is the perfect insurance of narrow banking and therefore the range where narrow banking dominates increases.

On the other hand fractional reserve banking has the advantage of partially compensating agents against the inflation tax on holding money because it pays interest on inside money $i_d > 0$ while under narrow banking interest on inside money is zero $i_d = 0$. This is the relative advantage of fractional reserve banking. It explains why the threshold $\tilde{\alpha}$ increases in the inflation tax γ/β. If the inflation tax is high holding money is very costly and thus the advantage of fractional reserve banking which provides a partial compensation for incurring this tax is very valuable. Thus the range where fractional reserve banking dominates increases in the inflation tax.

The following figure highlights this difference. In the narrow banking economy we have a full compensation against the inflation tax on the cash deposited by agents against non-monetary debt d' without compensation on inside money n. In the constrained fractional reserve banking economy we have a *partial* compensation of the inflation tax (because $1 + i_d$ is below the inflation tax) on the *full* stock of inside money n.

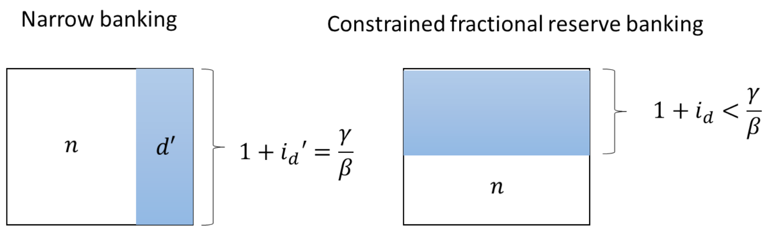


Figure 2.3: Interest under narrow and fractional reserve banking

In the fractional reserve banking economy from section 2.2 the relative advantage of narrow banking disappears. Banks now also offer non-monetary debt that perfectly compensates the sellers against the inflation tax as it also pays $1 + i'_d = \gamma/\beta$. With two types of debt also fractional reserve banks offer perfect insurance against liquidity risk and the threshold result disappears. Fractional reserve banking then strictly dominates narrow banking as proposition 4.1 shows.

2.4 Conclusion

The paper analyzed the welfare implications of narrow banking compared to the current fractional reserve banking system. Abstracting from fragility issues and focusing on the "monetary" role of banks where bank liabilities circulate as means of payment (inside money) the analysis showed that fractional reserve banking is beneficial because of the interest payments on inside money. Since inside money funds loans, it pays interest, compensating the agents for the inflation tax and thus reducing the welfare costs of inflation. The paper thus provides a more "monetary" argument for efficiency gains from fractional reserve banking which complements the classical analysis from Diamond and Dybvig [1983] where fractional reserve banking mobilizes investment in long-term, high-return assets. The paper also connects to the literature on the welfare costs of inflation. As observed by Lucas [2000] the possibility of demand deposits to pay interest should be taken into account when estimating the welfare costs of inflation using a measure like M1 which is a sum of non-interest bearing outside money and possibly interest-bearing inside money. The paper formalized this observation. It also showed that fractional reserve banking generally dominates narrow banking in terms of welfare because of these interest payments although narrow banking is modelled more carefully than in other papers. In this respect the paper also demonstrated how Berentsen et al. [2007] can be interpreted as a narrow banking economy where banks issue inside money and a non-monetary liability like long-term debt. Finally the paper analysed a situation where fractional reserve banking is constrained in the issuance of liabilities which could be interpreted as a distortion in the bank liability choice e.g. because inside money is subsidized by deposit insurance. The paper shows that only then and if banks are very constrained in their issuance of inside money narrow banking can yield higher welfare.

The broader message of the paper is that narrow banking systems in the spirit of the Chicago Plan where banks must back inside money fully with non-interest bearing outside money have efficiency costs in terms of foregone interest payments. Paying interest on outside money as proposed by Friedman [1960] would improve welfare but fractional reserve banking would still dominate narrow banking in such an environment. The interest rate on inside money would also incorporate the interest on outside money and still lie higher than the interest on outside money.

Various extensions could be addressed in further work. The model presented here is essentially a model of efficient liquidity provision and allocation. It could be augmented by banks having a role in capital accumulation and investment. It would also be interesting to quantify the welfare gains from inside money creation. Some quantitative estimates from New Monetarist models on the welfare costs of inflation are available. They could be complemented with a quantitative estimate of this model to provide a measure for the quantitative importance of these welfare gains which could then be set into relation to estimated costs of financial fragility. Another direction would be deviations from perfect competition. For example the discussion of the desirability of private seignorage by banks is impossible in a model where banks make zero profits. Finally a more complete analysis of the two banking systems should include financial fragility.

Bibliography

David Andolfatto. Reconciling Orthodox and Hererodox Views on Money and Banking. *working paper*, 2018.

David Andolfatto, Aleksander Berentsen, and Fernando M. Martin. Financial Fragility in Monetary Economics. *working paper*, 2016.

Pierpaolo Benigno and Roberto Robatto. Private money creation, liquidity crises, and government interventions. *Journal of Monetary Economics*, 106:42 – 58, 2019.

Aleksander Berentsen, Gabriele Camera, and Christopher Waller. Money, credit and banking. *Journal of Economic Theory*, 135(1):171 – 195, 2007.

Markus K. Brunnermeier and Dirk Niepelt. On the equivalence of private and public money. *Journal of Monetary Economics*, 106:27 – 41, 2019.

V.V. Chari and Christopher Phelan. On the social usefulness of fractional reserve banking. *Journal of Monetary Economics*, 65:1 – 13, 2014.

Douglas W. Diamond and Philip H. Dybvig. Bank Runs, Deposit Insurance, and Liquidity. *Journal of Political Economy*, 91(3):401 – 419, 1983.

Salomon Faure and Hans Gersbach. Money Creation in Different Architectures. *CEPR discussion paper*, 2019.

Milton Friedman. *A Program for Monetary Stability.* Fordham University Press, 1960.

Timothy Jackson and George Pennacchi. How Should Governments Create Liquidity? *working paper*, 2019.

Ricardo Lagos and Randall Wright. A unified framework for monetary theory and policy analysis. *Journal of Political Economy*, 113(3):463–484, 2005.

Robert E. Lucas, Jr. Inflation and welfare. *Econometrica*, 68(2):247–274, 2000.

George Pennacchi. Narrow banking. *Annual Review of Financial Economics*, 4(1):141–159, 2012.

Guillaume Rocheteau and Ed Nosal. *Money, Payments, and Liquidity.* MIT Press, 2017.

Jeremy C. Stein. Monetary policy as financial stability regulation. *The Quarterly Journal of Economics*, 127(1):57–95, 2012.

Stephen D. Williamson. Liquidity, monetary policy, and the financial crisis: A new monetarist approach. *American Economic Review*, 102(6):2570–2605, May 2012.

Appendix A

A.1 Economy with outside money and preference shock

In an economy without banks the value of an additional unit of cash for a buyer in $t+1$ is $\frac{u'(q_{b+1})}{p_{+1}}$ and for a seller just ϕ_{+1} since he cannot deposit and earn interest. Thus optimal cash holdings (2.7) for an agent in this economy solve:

$$\phi = \beta[(1-s)\frac{u'(q_{b+1})}{p_{+1}} + s\phi_{+1}].$$

Optimal production against cash is given by $c'(q_s) = p\phi$. Thus the stationarity equilibrium consumption in the first market without banks q_b solves:

$$\frac{u'(q_b)}{c'(\frac{1-s}{s}q_b)} = \frac{\gamma/\beta - s}{1-s}. \tag{A.1}$$

We see the RHS of equation (A.1) is equal to the RHS of A.2 if $s = 0$ (i.e. there is no preference shock) and that it is increasing in s. By the same logic as in A.1 we can thus conclude that if $s > 0$ the allocation is worse than in A.2 and the inefficiency increases in s.

A.2 Economy with outside money and no preference shock

Rocheteau and Nosal [2017] (p.138) show there exists a unique stationary equilibrium where all buyers and sellers have access to the first market (i.e. $\sigma = 1$ in their model) if $\gamma > \beta$ which solves:

$$\frac{u'(q_b)}{c'(q_s)} = \frac{\gamma}{\beta} \tag{A.2}$$

A.3 Use of inside and outside money in the first market

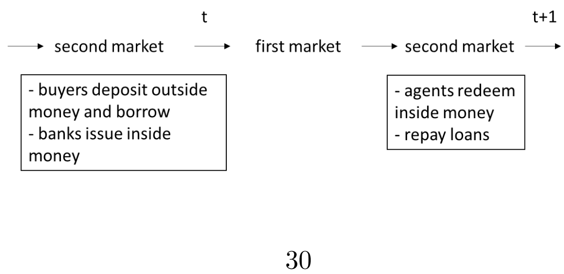
In this section we want to show that if both inside and outside money are used in the first market agents (weakly) prefer using inside money if the interest rate is non-negative. To make this point we will look at the choice of means of payments for buyers and neglect non-monetary debt and borrowing. Suppose a buyer arrives with m units of outside money. Still we denote the amount of outside money the buyer deposits in the bank for inside money as d_b and the amount of outside money she keeps as m'. Thus $m = d_b + m'$. To simplify we will also slightly change the interpretation of the price p in the first market. Before we assumed that it is expressed in terms of inside money and we formulated the means-of-payment-constraint as $pq_b \leq d_b + l_b$. Now we define p in terms of outside money in the next second market. Thus we can write: $pq_b \leq d_b(1 + i_d) + m'$. We will also just use outside money in the value function for the next second market. We can rewrite this modified buyer problem as:

$$V_b(m) = \max_{q_b, d_b, m'} \quad u(q_b) + W(d_b(1 + i_d) + m' - pq_b) \tag{A.3}$$
$$s.t. \quad pq_b \leq d_b(1 + i_d) + m'$$
$$d_b + m' = m$$

Using $m' = m - d_b$ in the problem we see that in both the right-hand side of the constraint and the amount of outside money holdings in the next second market you get the positive term $d_b i_d$. Thus the marginal benefits of depositing are positive if $i_d > 0$ and a buyer would like to set d_b as high as possible, i.e. $d_b = m$ and $m' = 0$. This illustrates that a buyer strictly prefers to use inside money if the interest on inside money is positive.

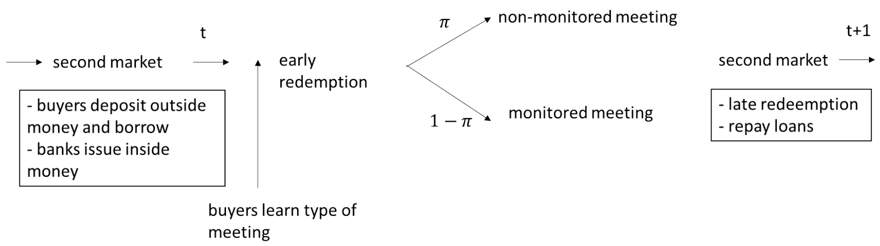
A.4 A model with early redemptions

So far the need for banks to hold outside money was motivated by the assumption that clearing and settling the transactions with inside money in the first market takes a minimal amount of outside money proportional to the inside money used (α). However, the concrete clearing/settlement process was not modelled. Building on Williamson [2012] I now motivate the outside money holdings of banks differently. Compared to the baseline model I will make some simplifying assumptions. First I assume there is a unit mass of buyers and sellers each and there is no preference shock. This limits the role of banks to the provision of liquidity and means only buyers will acquire money. I will also ignore the possibility that banks issue non-monetary debt and buyers will already acquire inside money by depositing and borrowing in the second market. Without preference shock this does not change the problem. Compared to figure 2.1 the sequence of events then simplifies to:



Following Williamson [2012] I will also depart from perfect competition in the first market and assume bilateral meetings between buyers and sellers where buyers make take-it-or-leave-it (TIOLI) offers to sellers. And I will assume linear disutility of working in the first market, $c(q) = q$, for sellers and CRRA-utility for buyers with risk aversion below 1, $u(q) = \frac{q^{1-\sigma}}{1-\sigma}$ with $\sigma < 1$ for buyers.[1]

The crucial novelty is that now there are two types of meetings in the first market for buyers. With probability π buyers go to a *non-monitored* meeting where they can only use outside money. And with probability $1 - \pi$ they go to a *monitored* meeting where they can use inside money like in the baseline model (if $\pi = 0$ the new model is a simplified version of the baseline model). Buyers acquire inside money by depositing cash and borrowing in the second market. But since they may need cash if they go to a non-monitored meeting the bank allows for *early* redemptions. After knowing the type of meetings buyers can redeem inside money into outside money at interest rate i_{d1} before the first market. Inside money which is not redeemed early will be redeemed in the following second market at the interest rate i_{d2} as in the baseline model. The sequence of events is then as follows:



Given the simplifications we can focus on the buyers. The buyers acquire inside money by depositing outside money, d, and borrowing l. The total amount of inside money a buyer acquires in a second market is $n = l + d$. Let the amount consumed in a non-monitored meeting be q^c (for "cash") and q in a monitored meeting. Also let n^c be the amount of inside money a buyer redeems before a non-monitored meeting and n' be the amount of inside money used in a monitored meeting where $n^c, n' \leq n$ for feasibility. In a non-monitored meeting the buyer redeems n^c units of inside money for cash at rate $1 + i_{d1}$. The value of this outside money is $\phi n^c(1 + i_{d1})$ for a seller in the next second market and since we assume take-it-or-leave-it offers by buyers this is exactly the amount produced, i.e. $q^c = \phi n^c(1 + i_{d1})$. In a monitored meeting the buyer uses n' units of inside money to pay the seller where by the same reasoning we must have that $q = \phi n'(1 + i_{d2})$. The problem of a representative buyer can then be written as:

[1] These assumptions do not fundamentally change the results. Assuming bilateral meetings with TIOLI offers and linear seller utility in the baseline model yields the same allocation $u'(q) = \frac{1+i}{1+i_d}$ as in the competitive model, see (2.27).

$$\max_{d,l,n^c \leq n, n' \leq n} \quad -\phi_{-1}d + \beta\pi[u(\phi n^c(1+i_{d1})) + \phi(n-n^c)(1+i_{d2})] \tag{A.4}$$
$$\beta(1-\pi)[u(\phi n'(1+i_{d2})) + \phi(n-n')(1+i_{d2})]$$
$$-\beta\phi l(1+i)$$

We can distinguish four possible cases solving this problem:

a) costless inside money ($i_{d1} < i_{d2} = i$): In this case inside money is costless to hold and the buyer holds so much that he is unconstrained in both types of meetings. The unconstrained consumption levels are:

$$u'(\tilde{q}^c) = \frac{1+i_{d2}}{1+i_{d1}} \quad , \quad u'(q^*) = 1 \tag{A.5}$$

i.e. in the non-monitored meeting the buyer wants to consume $\tilde{q}^c$ and in the monitored meeting he consumes the first best quantity q^*. The return on inside money equals the inflation tax $1 + i_{d2} = 1 + i = \frac{\phi}{\beta\phi_{+1}}$ and the buyer doesn't use all his inside money in both types of meetings ($n^c, n' < n$). The amount of real inside money holdings ϕn is undetermined when inside money is costless to hold. To be unconstrained in the monitored meeting a buyer needs at least $q^*/(1+i_{d2}) = n^*$ units of real inside money and to be unconstrained in the non-monitored meeting the buyer needs at least $\tilde{q}^c/(1+i_{d1}) = \tilde{n}$ units. With $\sigma < 1$ we have $\tilde{n} < n^*$. Thus buyers need less real inside money to be unconstrained in the non-monitored meeting and real inside money holdings ϕn must lie in (n^*, ∞). Defining $r = n^c/n$ as the *redemption rate* in the first market, i.e. the fraction of inside money redeemed by buyers in the non-monitored meeting, we must have also have $r \in \left(\frac{\tilde{n}}{n^*}, 0\right)$.

b) medium return on inside money ($i_{d1} < \tilde{i}_{d2} \leq i_{d2} < i$): In this case inside money is costly to hold ($i_{d2} < i$) but the return is still high enough that buyers hold enough real inside money to be unconstrained in the non-monitored meetings, i.e. $q^c = \tilde{q}^c$ and $n^c < n$, but they are constrained in the monitored ones ($n' = n$) and $q = \phi n(1 + i_{d2})$. The consumption levels are:

$$u'(\tilde{q}^c) = \frac{1+i_{d2}}{1+i_{d1}} \quad , \quad u'(q) = \frac{1+i-\pi(1+i_{d2})}{(1-\pi)(1+i_{d2})} \tag{A.6}$$

and real inside money holdings solve:

$$\phi n = \Big(\frac{(1-\pi)(1+i_{d2})}{1+i-\pi(1+i_{d2})}\Big)^{1/\sigma} \frac{1}{1+i_{d2}} \tag{A.7}$$

Note that real inside money holdings increase in the return on inside money i_{d2} for $\sigma < 1$. The lower bound for the return on inside money, $\tilde{i}_{d2}$, is the value for i_{d2} where the buyer holds just enough real inside money to be unconstrained in non-monitored meetings, i.e. to consume $\tilde{q}^c$. Thus the threshold solves:

$$\tilde{n}(1 + i_{d1}) = \tilde{q} \tag{A.8}$$

and we can express this solution also in terms of real inside money holdings, $\phi n \in [\tilde{n}, n^*)$.

c) low return on inside money ($i_{d1} < i_{d2} < \tilde{i}_{d2} < i$): In this case the return on inside money is so low that buyers are constrained in both types of meetings ($n^c = n' = n$). The consumption levels and real inside money holdings solve:

$$\pi u'(q^c)(1 + i_{d1}) + (1 - \pi)u'(q)(1 + i_{d2}) = 1 + i \tag{A.9}$$

$$q^c = \phi n(1 + i_{d1}) \quad , \quad q = \phi n(1 + i_{d2})$$

and

$$\phi n = \Big(\frac{\pi(1 + i_{d1})^{1-\sigma} + (1 - \pi)(1 + i_{d2})^{1-\sigma}}{1 + i}\Big)^{1/\sigma} \tag{A.10}$$

This solution occurs if real inside money holdings are below the amount to consume $\tilde{q}^c$ in non-monitored meetings, i.e. $\phi n < \tilde{n}$

d) low return on inside money without spread ($i_{d1} = i_{d2} = i_d < i$): In this case inside money is costly to hold and its return is the same when redeemed early or late.[2] This implies buyers are constrained in both types of meetings and consumption in both meetings is the same solving

$$u'(q) = \frac{1 + i}{1 + i_d} \tag{A.11}$$

with real inside money holdings

$$\phi n = \Big(\frac{1 + i_d}{1 + i}\Big)^{1/\sigma} \frac{1}{1 + i_d} \tag{A.12}$$

In any of these cases the indifference condition between the two ways of acquiring inside money (borrowing and depositing cash) must hold as in the baseline model (see (2.23):

$$\frac{\phi_{-1}}{\phi}\frac{1}{\beta} = 1 + i \tag{A.13}$$

Also, the following relationship between the interest rates must hold in any equilibrium:

$$0 \leq i_{d1} \leq i_{d2} \leq i \tag{A.14}$$

The inequalities ensure that buyers prefer using inside money, that only buyers going to a non-monitored meeting redeem early[3]. and that the solution for real inside money holdings is bounded.

[2]We abstract from the case when $i_{d1} = i_{d2} = i$ which can never be feasible for the bank as we see below.

[3]We abstract from the possibility of belief-driven redemptions in the spirit of Diamond and Dybvig [1983]

bank problem: The bank maximizes total cash profits taking the redemption rate r as given. Total redemptions before the first market are given by $\pi rn(1+i_{d1})$[4] and this needs to be smaller or equal to d, the total amount of cash deposits. This constraint is similar to the reserve constraint in the model before. I also introduce a quantity constraint. I assume the monitoring/enforcement technology of banks for loans is imperfect and loans cannot be bigger than a threshold value $\bar{l}$.[5] The problem of the bank is:

$$\max_{l,d} = d - \pi rn(1+i_{d1}) + l(1+i) - (n - \pi rn)(1+i_{d2})$$

$$s.t. \quad \pi rn(1+i_{d1}) \leq d, \quad l \leq \bar{l}$$

We see redemptions modify the bank problem in two ways (compare it with (2.2)). It decreases outside money holdings and it decreases the outstanding inside money of banks. Thus it decreases both assets and liabilities. We can reformulate the objective function as follows:

$$l(i - \pi ri_{d1} + (1-\pi r)i_{d2}) - d(\pi ri_{d1} + (1-\pi r)i_{d2})$$

Thus if $0 < \pi ri_{d1} + (1-\pi r)i_{d2} < i$ the bank wants to set loans as high as possible and cash deposits as low as possible. This implies both constraints should bind and we get:

$$\bar{l} = \frac{1-\pi r(1+i_{d1})}{\pi r(1+i_{d1})}d \tag{A.15}$$

and the zero profit condition implies:

$$1+i_{d2} = \frac{1-\pi r(1+i_{d1})}{1-\pi r}(1+i) \tag{A.16}$$

Now consider the case when the first constraint does not bind. In this case the bank holds more outside money than she needs for redemptions in the non-monitored meetings. We know if $\pi ri_{d1} + (1-\pi r)i_{d2} > 0$ the bank wants to set d as low as possible because she makes losses by holding outside money. Thus the bank will only accept excess reserves if $\pi ri_{d1} + (1-\pi r)i_{d2} = 0$, i.e. holding outside money has no costs. Then, zero profits would require the loan rate to come down to the same level, i.e. $i = 0$. But since the indifference condition (A.13) must hold, this can never be an equilibrium. If the loan rate lies below the inflation tax, agents would not be willing to deposit

[4]We consider a large bank with lots of buyers where the fraction of buyers going to a non-monitored meeting is approximately π.

[5]This is to make the model comparable to Williamson [2012]. In Williamsons model buyers deposit goods in banks and banks invest them into outside money and nominal government bonds. The government bonds can be used as means of payment in monitored meetings (they serve a very similar role to inside money in my model). In his model there is also a quantity constraint which steers the aggregate issuance of government bonds, $\delta = \frac{M}{M+B}$.

outside money in the bank, they would only borrow. But if $i > 0$ and $\pi r i_{d1} + (1-\pi r) i_{d2} = 0$ banks make positive profits in equilibrium. One can avoid this problem by assuming that depositing and borrowing are coupled, i.e. buyers always deposit and borrow at the same bank. Then it makes sense for banks to compete for new depositors/borrowers in a situation where $i > 0$ and $\pi r i_{d1} + (1 - \pi r) i_{d2} = 0$ and thus $\pi r i_{d1} + (1 - \pi r) i_{d2}$ ultimately increases to a level where banks again make zero profits. This is achieved when

$$(i - \pi r i_{d1} + (1 - \pi r) i_{d2}) \bar{l} = (\pi r i_{d1} + (1 - \pi r) i_{d2}) d \tag{A.17}$$

Thus the equilibrium conditions for banks with excess reserves are given by $l = \bar{l}$ (given $i > \pi r i_{d1} + (1 - \pi r) i_{d2}$ still holds), (A.17) and $d > \pi r n(1 + i_{d1})$ which implies

$$1 + i_{d2} < \frac{1 - \pi r(1 + i_{d1})}{1 - \pi r}(1 + i) \tag{A.18}$$

stationary equilibrium: As usual stationarity implies $\gamma = \phi/\phi_{+1}$ and thus the loan rate must equal the inflation tax or $1 + i = \gamma/\beta$ from (A.13) as in the baseline model. Also market clearing for outside money implies $d = M_{-1}$. Thus the aggregate ratio of outside to inside money under a binding borrowing constraint is fixed and we can define this ratio as $\bar{\alpha}$[6]

$$\bar{\alpha} = \frac{M_{-1}}{M_{-1} + \bar{l}}$$

We can rewrite equilibrium interest rates without excess reserves using (A.15) and (A.16) as

$$\begin{aligned} 1 + i_{d1} &= \frac{\bar{\alpha}}{\pi r} \\ 1 + i_{d2} &= \frac{(1 - \bar{\alpha})(1 + i)}{1 - \pi r} \end{aligned} \tag{A.19}$$

Using the conditions on the interest rates, (A.14), this yields a feasible range for the redemption rate $r \in (\frac{\bar{\alpha}}{\pi((1-\alpha)\gamma/\beta+\alpha)}, \bar{\alpha}/\pi)$

With excess reserves using (A.17) we get:

$$\pi r(1 + i_{d1}) + (1 - \pi r)(1 + i_{d2}) = \bar{\alpha} + (1 - \bar{\alpha})(1 + i) \tag{A.20}$$

and we need $\bar{\alpha} > \pi r(1 + i_{d1})$.

We can now characterize the four equilibria:

a) equilibrium with costless and plentiful inside money: In this equilibrium inside money is costless to hold ($i_{d2} = i = \gamma/\beta - 1$). Therefore it must be from (A.19) that $r = \bar{\alpha}/\pi$ and $i_{d1} = 0$, i.e. the

[6]Note the inverse relation between borrowing constraints and $\bar{\alpha}$: if the borrowing constraints are tight ($\bar{l}$ is low) $\bar{\alpha}$ is high and the other way round.

interest rate for early redemptions is zero. Given these interest rates equilibrium consumption in both types of meetings solves:

$$u'(\tilde{q}^c) = \gamma/\beta \quad , \quad u'(q^*) = 1$$

Since $\tilde{q}^c = \phi rn$ with $i_{d1} = 0$ and r is pinned down, in equilibrium real inside money holdings must be $\phi n = \tilde{q}^c/r$. We also know real inside money holdings ϕn must be at least $\frac{q^*}{\gamma/\beta}$. Thus for existence of this equilibrium $\bar{\alpha}$ cannot be higher than some threshold α^* and the equilibrium exists for $\bar{\alpha} \in (0, \alpha^*)$:

$$\alpha^* = \frac{\pi\tilde{q}^c(\gamma/\beta)}{q^*} < \pi \tag{A.21}$$

In this equilibrium inside money is plentiful. The lending technology of the bank is very good ($\bar{l}$ is high and $\bar{\alpha}$ is low) so banks can issue a lot of inside money which allows them to pay a return on inside money which makes it costless to hold. In fact, there are two forces. Banks need to be able to pay an interest rate on inside money equal to the loan rate (inflation tax). With a zero interest rate on early redemptions this implies that after the early redemptions the bank has the same amount of inside money and loans outstanding (and thus under zero profits interest rates on both sides of the balance sheet are equalized). In an aggregate sense we then need $(1-\pi r)(M_{-1}+\bar{l})$ (the amount of inside money outstanding after redemptions) equals the amount of loans outstanding $\bar{l}$. This is only feasible if $r = \bar{\alpha}/\pi$. The redemption rate needs to have a certain size and for example if $\bar{\alpha} > \pi$ it is never feasible for the bank to pay $i_{d2} = i$. The second condition is that buyers need to have enough real inside money holdings for unconstrained consumption in the monitored market. In principle if holding inside money is costless buyers can acquire as much as they want. However, the higher ϕn the lower the redemption rate needs to be because consumption in the non-monitored market solves $\phi n^c = \phi rn = \tilde{q}^c$ and at some point r would be below $\bar{\alpha}/\pi$ even if real inside money holdings are just enough that the buyer is unconstrained in the monitored meeting.

Note that in this equilibrium inside money is neutral. $\bar{\alpha}$ has no effect on consumption in the two types of meetings as long as it satisfies (A.21). So the quantity theory holds. The allocation is not neutral to changes in the inflation tax however. A higher inflation tax γ/β decreases real outside money holdings of banks and thus consumption in non-monitored meetings. A higher inflation tax also decreases α^* and thus narrows the range of the equilibrium. This equilibrium is equivalent to the plentiful interest bearing asset case in Williamson [2012].

b) equilibrium with medium return on inside money: In this equilibrium buyers are unconstrained in the non-monitored meeting because the return on inside money is quite high ($\tilde{i}_{d2} < i_{d2} < i$). So they prefer keeping some inside money and redeeming it in the second market only. This also implies the redemption rate r is below 1. In the monitored meetings buyers are still constrained because $i_{d2} < i$ and inside money is costly to hold. This equilibrium is pinned down by six equations in $(q, \tilde{q}^c, i_{d1}, i_{d2}, r, \phi n)$: (A.6), (A.7) and (A.19) and $\tilde{q}^c = \phi nr(1 + i_{d1})$.

We know this equilibrium exists if $i_{d2} \geq \tilde{i}_{d2}$ pinned down by (A.8). We also know at $\tilde{i}_{d2}$ the redemption rate is 1 and thus the interest rates are fully pinned down by $\bar{\alpha}$ and π. We can thus define a threshold $\tilde{\alpha}$ which solves (A.8) for interest rates satisfying (A.19) with $r = 1$ and formulate $i_{d2} \geq \tilde{i}_{d2}$ equivalently as $\alpha \leq \tilde{\alpha}$. Rewriting (A.8) we get:

$$\left(\frac{1 + i_{d2}}{1 + i_{d1}}\right)^{1-\sigma} = \frac{1 + i - \pi(1 + i_{d2})}{(1 - \pi)(1 + i_{d2})} \tag{A.22}$$

Suppose $\tilde{\alpha}$ is π. From (A.19) we would then get $i_{d2} = i$ and $i_{d1} = 0$. Clearly this violates (A.22). We need $\tilde{i}_{d2} < i$ and ergo $\tilde{\alpha} > \pi$. Now suppose $\tilde{\alpha} = \underline{\alpha}$ where

$$\underline{\alpha} = \frac{\pi(\gamma/\beta)}{1 - \pi + \pi(\gamma/\beta)} \tag{A.23}$$

is the value where both interest rates are equal ($i_{d2} = i_{d1} = i_d$), according to (A.19) with $r = 1$. Also then (A.22) is violated and we need $\tilde{\alpha} < \underline{\alpha}$ or $\tilde{i}_{d2} > i_d$. Thus we must have $\tilde{\alpha} \in (\underline{\alpha}, \pi)$ and since $\alpha^* < \pi$ from (A.21) we must have $\tilde{\alpha} < \alpha^*$ and the interval $(\alpha^*, \tilde{\alpha})$ is non-empty.

What happens to expected welfare? Denoting the optimal choice of real inside money holdings as n^* and the optimal redemption rate as r^* we can rewrite (A.4):

$$\begin{aligned} U_b(n^*, r^*) = -n^*(1 + i) + \pi u(n^* r^*(1 + i_{d1})) + \pi n^*(1 - r^*)(1 + i_{d2}) \\ +(1 - \pi)u(n^*(1 + i_{d2})) \end{aligned}$$

Using (A.16) to substitute for $1 + i_{d1}$ and applying the envelope theorem:

$$\frac{dU_b(n^*, r^*)}{d(1 + i_{d2})} > 0 \quad \text{if} \quad \frac{(1 + i)(1 + i_{d1})}{1 + i_{d2}} > \frac{1 - \pi r^*}{(1 + i)/(1 + i_{d2}) - \pi r^*}$$

which holds for $i > i_{d2}, i_{d1} > 0$. Thus as $\bar{\alpha}$ decreases from $\tilde{\alpha}$ to α^* and i_{d2} increases from $\tilde{i}_{d2}$ to i expected welfare increases (at $i_{d2} = i$ and $i_{d1} = 0$ the derivative is zero). Although buyers in non-monitored meetings loose and buyers in monitored meetings gain both gain in expected terms. And more inside money (lower $\bar{\alpha}$) is beneficial in this equilibrium.[7]

c) equilibrium with low return and scarce inside money: In this equilibrium $i_{d1} < i_{d2} < \tilde{i}_{d2} < i$ and buyers are constrained in both types of meetings. The equilibrium is pinned down by (A.9), (A.10) and (A.19) with $r = 1$. The conditions on the interest rates then tell us we need $\bar{\alpha} < \tilde{\alpha}$ and $\bar{\alpha} > \underline{\alpha}$. Thus it exists in the range $\bar{\alpha} \in (\underline{\alpha}, \tilde{\alpha})$.

[7] This equilibrium does not exist in Williamson [2012] since he does not consider inside money and redemptions explicitly.

Using (A.10) one can show that $\frac{d(\phi n)}{d\bar{\alpha}} < 0$ for $\bar{\alpha} \in (\underline{\alpha}, \tilde{\alpha})$ and thus $\frac{dq}{d\alpha} < 0$. Thus if we decrease $\bar{\alpha}$ from $\underline{\alpha}$ to $\tilde{\alpha}$ real inside money holdings and consumption in the monitored meeting increase. Although a higher $\bar{\alpha}$ increases inside money holdings through i_{d1}, this is are overcompensated by the decrease in i_{d2}. The effect on consumption in the non-monitored meeting is not clear and depends on the level of the inflation tax. If the inflation tax is not too high[8] a lower $\bar{\alpha}$ decreases consumption in the non-monitored market. In this case the decreasing effects from the lower interest rate i_{d1} overcompensate the increasing effects from higher real inside money holdings.

Again denoting optimal real inside money holdings as n^* indirect buyer utility can be written as:

$$U_c(n^*) = -n^*(1+i) + \pi u(n^*(1+i_{d1})) + (1-\pi)u(n^*(1+i_{d2}))$$

Using (A.16) to substitute for $1+i_{d1}$ and applying the envelope theorem we see that:

$$\frac{dU_c(n^*)}{d(1+i_{d2})} > 0 \quad \text{if} \quad \frac{u'(q^c)}{u'(q)} < 1+i \quad \text{and} \quad \frac{1+i_{d2}}{1+i_{d1}} < (1+i)^{1/\sigma}$$

which must hold for $i_{d2} < i$. As we decrease $\bar{\alpha}$ (increase i_{d2}) expected welfare increases in this equilibrium (although also here buyers in the non-monitored meetings might loose). This also implies that equilibrium *b*) dominates equilibrium *c*) in terms of welfare. This equilibrium is analogue to the equilibrium with scarce interest bearing assets in Williamson [2012].

d) equilibrium with excess reserves: In this equilibrium $i_{d1} = i_{d2} = i_d$ and buyers are constrained in both markets. We know interest rates are equal at $\bar{\alpha} = \underline{\alpha}$. If $\bar{\alpha} > \underline{\alpha}$ and (A.19) holds we would get $i_{d1} > i_{d2}$ which violates our assumptions on the interest rates. Thus when $\bar{\alpha} > \underline{\alpha}$ banks cannot pay out all outside money in the non-monitored meeting and we must have excess reserves (the first constraint of the bank must be slack). Using $i_{d1} = i_{d2} = i_d$ on (A.20):

$$1 + i_d = \bar{\alpha} + (1-\bar{\alpha})(1+i) \tag{A.24}$$

Thus this equilibrium solves (A.11) and (A.24) and $\phi n = q/(1+i_d)$ in $\bar{\alpha} \in (\underline{\alpha}, 1)$. Note that (A.24) is exactly the same expression we got in the basic model (2.4). The interest rate on inside money is a weighted average of the return on outside money and the return on loans. Since consumption in monitored and non-monitored meetings is also identical to the basic model we get exactly the same expression as (2.27) (with a linear disutility of working for sellers, i.e. $c'(q_s) = 1$)

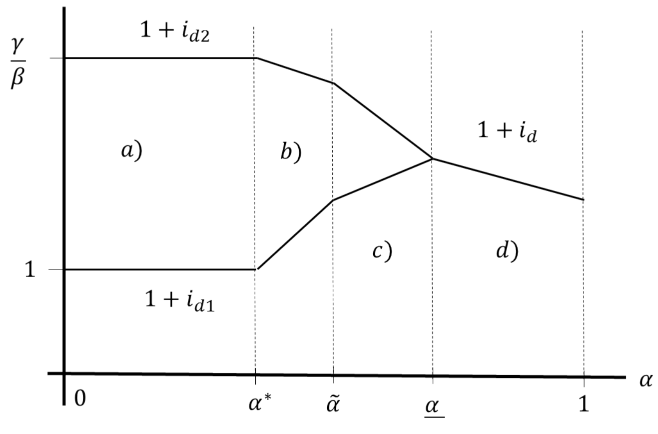
$$u'(q) = \frac{1+i}{1+i_d} = \frac{\gamma/\beta}{(1-\alpha)\gamma/\beta + \alpha} \tag{A.25}$$

The welfare implications are also identical to the basic model. The allocation is non-neutral with

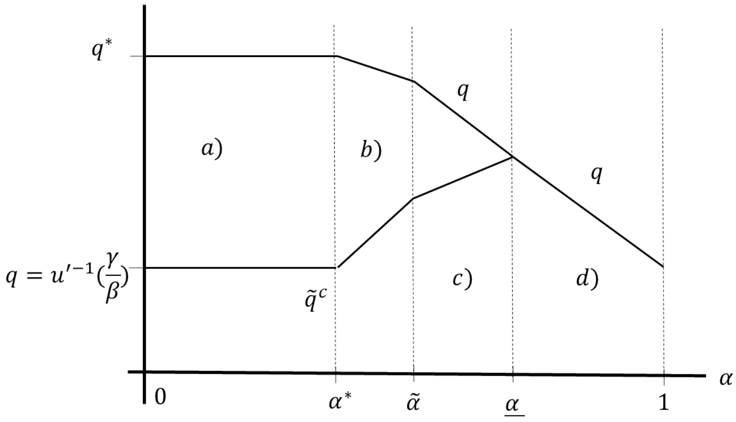
[8] Concretely if $\gamma/\beta < \frac{\pi+(1-\pi)\sigma}{\pi(1-\sigma)}$.

respect to $\bar{\alpha}$. The higher $\bar{\alpha}$ (the less inside money) the lower welfare.[9] Finally note that a as in the basic model a narrow banking system would be when $\bar{\alpha} = 1$ or $\bar{l} = 0$ i.e. banks can do no lending. As in the basic model outside and inside money are perfect substitutes and the question whether buyers will go to a monitored or a non-monitored meeting is irrelevant because both means of payment are equivalent. The results from the basic models in terms of welfare also apply here. Fractional reserve banking with $\bar{\alpha} < 1$ dominates narrow banking where consumption in both types of meetings solves $u'(q) = \gamma/\beta$.

The following figures summarize the four equilibria. The more elaborate model repeats and refines the basic message from the sections before. As in the basic model in equilibria $b), c), d)$ more inside money creation (a lower $\bar{\alpha}$) is beneficial because it increases the return on inside money which compensates agents for the inflation tax. However, and this is the first refinement, in equilibria $b)$ and $c)$ more inside money creates winners and losers. Although in expected terms buyers are better of, they may lose in the non-monitored meetings. Marginal utility is higher when interest payments are concentrated on the monitored meetings and thus i_{d2} increases and interest payments in the non-monitored meetings (i_{d1}) shrinks. The second refinement is that in equilibrium $a)$ when inside money creation is sufficiently easy ($\bar{\alpha} < \alpha^*$) inside money has no real effects anymore. Thus we loose the non-neutrality result. The refined model offers also a reinterpretation of the basic model as a situation where banks are very constrained in their lending and are forced to hold excess reserves in equilibrium.



[9]These results are in contrast to Williamson [2012] who gets a "liquidity trap" equilibrium in the analogue situation where the rate of return on outside money and bonds are equal. In his model the return on bonds is not linked to the inflation tax as the loan rate is here over the indifference condition (A.13). Thus if bonds are very scarce, their return would be below the return on outside money if only bonds were used in the monitored meeting. Therefore agents want to use outside money in both meetings and the returns of bonds and outside money and the consumption levels must be equalized. But this implies $\bar{\alpha}$ has no real effects in this equilibrium in his model.


q^*
q
$a)$
$b)$
q
$q = u'^{-1}(\frac{\gamma}{\beta})$
$c)$
$d)$
$\tilde{q}^c$
α
0
α^*
$\tilde{\alpha}$
$\underline{\alpha}$
1

Appendix B

B.1 proof of proposition 4.1

Proof. To derive (2.27) we conjectured that $i = i_d' > i_d > 0$ in equilibrium. Note that if $\alpha \in (0,1)$ and $\gamma > \beta$ equilibrium interest rates given by (2.24), (2.25) and (2.26) satisfy this. Given that $i > i_d$ the buyer will always use all his money in the first market, $\lambda = \phi(i - i_d) > 0$. Then (2.11) and (2.6) are strictly concave in q_b, l_b, m and the first order conditions are sufficient for a unique maximum. Market clearing in the first market (2.21) pins down a unique q_s even if $c(q_s)$ is not strictly convex (and (2.16) strictly concave). Thus the solution to (2.27) must be unique.

Comparative statics: Differentiate the left-hand side of (2.27), $\frac{u'(q_b)}{c'(q_s)}$, with respect to q_b

$$\frac{\partial(\frac{u'(q_b)}{c'(q_s)})}{\partial q_b} = \frac{u''(q_b)c'(q_s) - u'(q_b)c''(q_s)\frac{1-s}{s}}{c'(q_s)^2} < 0 \tag{B.1}$$

From the strict concavity of $u(q)$, $u''(q) < 0$ and therefore (B.1) must decrease in q_b. At the first best allocation q^* (2.1) $\frac{u'(q^*)}{c'(q_s)} = 1$ and at any q_b solving (2.27) $\frac{u'(q_b)}{c'(q_s)} = \frac{1+i}{1+i_d} > 1$ since $i > i_d$ for $\alpha \in (0,1)$ and $\gamma > \beta$. Thus $\frac{u'(q^*)}{c'(q_s)} < \frac{u'(q_b)}{c'(q_s)}$ and therefore $q_b < q^*$ from (B.1). Thus we showed that for any $i > i_d$ the allocation will be inefficient. By the same argument any change *increasing* the right-hand side of (2.27) higher above 1 will decrease q_b further from q^*. Since $\frac{1+i}{1+i_d}$ increases in α we must have $\frac{\partial q_b}{\partial \alpha} < 0$ i.e. equilibrium consumption and welfare decreases in α. Also since $\frac{1+i}{1+i_d}$ increases in γ/β we must have $\frac{\partial q_b}{\partial \gamma/\beta} < 0$ i.e. equilibrium consumption and welfare decreases in the inflation tax.

iii)tov): The right-hand side of (2.27) $\gamma/\beta\frac{1}{1+i_d}$ is strictly below the right-hand side of an economy without banks (A.1) $\frac{\gamma/\beta - s}{1-s}$. Therefore by the same argument as above q_b (and welfare) are higher under fractional reserve banking than without banks. The results for *iv*) and *v*) are simply obtained by sticking in the limit values of α $(0,1)$ into the right-hand side of 2.27. At $\alpha = 1$ the right-hand sides of (A.2) in appendix A.2 and in (2.27) coincide.

■

B.2 proof of proposition 4.2

Proof. If $\alpha \in (0,1)$ and $\gamma > \beta$ equilibrium interest rates under constrained fractional reserve banking (2.37) and (2.36) satisfy $i > i_d > 0$. Then the same steps as in the first proof apply.

To verify *ii*) we use the same logic as in the first proof. To show that q_b and welfare in (2.38) is below first best q^* it is sufficient to show that the right-hand side of (2.38) is above 1. This holds if $\alpha \in (0,1)$ and $\gamma > \beta$. The comparative statics follow the same logic as above. If the partial derivative of the right-hand side of (2.38) with respect to the parameters s, α or γ/β increases the interest rate spread $\frac{1+i}{1+i_d}$ resp. the right-hand side of (2.38), q_b and welfare decreases. We have:

$$\frac{\partial \frac{1+i}{1+i_d}}{\partial \gamma/\beta} = \frac{\alpha(1-\alpha s)}{((1-\alpha)\gamma/\beta + \alpha(1-s))^2} > 0$$
$$\frac{\partial \frac{1+i}{1+i_d}}{\partial \alpha} = \frac{(1-s)^2\alpha + \gamma/\beta(\gamma/\beta - 1))}{((1-\alpha)\gamma/\beta + \alpha(1-s))^2} > 0$$
$$\frac{\partial \frac{1+i}{1+i_d}}{\partial s} = \frac{\gamma/\beta(1-\alpha) + \alpha^2(\gamma/\beta - 1))}{((1-\alpha)\gamma/\beta + \alpha(1-s))^2} > 0$$

Thus indeed $\frac{\partial q_b}{\partial \gamma/\beta} < 0$, $\frac{\partial q_b}{\partial \alpha} < 0$ and $\frac{\partial q_b}{\partial s} < 0$.

iii): As s goes to zero q_b given by (2.38) converges to the allocation without preference shock (2.27). Since increasing s decreases q_b from *ii*) q_b and welfare under unconstrained fractional reserve banking (2.27) must be higher.

iv): Compare the right-hand side of the no-bank equilibrium (A.1) with the right-hand side of (2.38) to see that the former is strictly higher for $\alpha \in (0,1)$ and identical for $\alpha = 1$.

v): Finding $\tilde{\alpha}$ requires equalizing the right-hand sides of the narrow banking equilibrium allocation (A.2), see proposition 4.1 *iv*) and the constrained fractional reserve banking equilibrium (2.38):

$$\frac{\gamma}{\beta} = \frac{\gamma/\beta - \tilde{\alpha}s}{(1-\tilde{\alpha})\gamma/\beta + \tilde{\alpha}(1-s)}$$

which yields

$$\tilde{\alpha} = \frac{\gamma/\beta}{\gamma/\beta + s}$$

For *vi*) use $\alpha = 0$ in the right-hand side of (2.38) which then yields q^* from (2.1).

■

Liquidity, the Mundell-Tobin Effect, and the Friedman Rule[1]

When it comes to optimal monetary policy, there is a stark contrast between central bankers and monetary theorists. Most central banks in developed economies follow an inflation target of around 2% annually, and there is a general agreement among central bankers that deflation has to be avoided. Meanwhile, most theoretical models find that the Friedman rule, i.e., setting the inflation rate such that the opportunity cost of holding money balances is zero, is the optimal monetary policy. Since zero opportunity costs for holding money implies deflation in standard models, this prediction clearly differs from what central bankers believe to be optimal. The Friedman rule has been found to be optimal by Friedman himself in a model with money in the utility [Friedman, 1969], but also in a variety of other monetary models such as cash-in-advance [Grandmont and Younes, 1973, Lucas and Stokey, 1987], spatial separation [Townsend, 1980], and New Monetarism [Lagos and Wright, 2005].[2] While there have been alterations of these models that render the Friedman rule suboptimal, these usually rely on additional frictions that do not seem to be major concerns for central bankers in reality.[3]

One potential reason to run inflation above the Friedman rule is to stimulate investment. This mechanism is captured by the Mundell-Tobin effect (Mundell [1963] and Tobin [1965]). The Mundell-Tobin effect predicts that an increase in the return on nominal assets such as bonds or fiat money (e.g., a reduction in inflation) crowds out capital investment. However, inflation above the Friedman rule reduces people's willingness to hold liquid assets. If certain trades can only be settled with liquid assets, higher inflation rates thus reduce quantities traded. This implies that there is a tradeoff between the benefits of a high return on liquid assets and the costs associated with reduced capital investment due to the Mundell-Tobin effect. We investigate this tradeoff in a model that combines the overlapping generations (OLG) framework a la Wallace [1980] with a New Monetarist model a la Lagos and Wright [2005] (LW). This approach allows us to find novel results regarding both of these literatures, and settle some debates, as we explain in the literature review below. In particular, we find that the Friedman rule is optimal in our model if and only if there is no Mundell-Tobin effect. If there is a Mundell-Tobin effect at the Friedman rule, an increase in inflation leads to a first-order welfare gain from reducing hours worked, and only to a

[1]Co-authored with Lukas Altermatt, University of Essex, UK.

[2]See Schmitt-Grohé and Uribe [2010] for an overview.

[3]E.g. for the New Monetarist literature: theft as in Sanches and Williamson [2010], incomplete tax instruments as in Aruoba and Chugh [2010], or socially undesirable activities financed by cash as in Williamson [2012].

second-order welfare loss from reducing consumption in markets where liquidity matters. Further, we can show that a Mundell-Tobin effect is more likely to occur at the Friedman rule if agents care less about consumption smoothing, and if capital is relatively liquid.

In our model, each period is divided into two subperiods, called CM and DM. Agents are born at the beginning of the CM and live until the end of the CM of the following period; i.e., they are alive for three subperiods. There are two assets in the economy, productive capital and fiat money. When agents are born, they immediately learn whether they will be a buyer or a seller during the DM. During the first CM of their lives, all agents can work at linear disutility and accumulate capital and fiat money. In the DM, sellers can work at linear disutility and produce a DM good. Buyers cannot work during the DM, but they get concave utility from consuming the DM good. With some probability, buyers are relocated during the DM. If they are relocated, they can only use fiat money to settle trades, because we assume that capital is immobile and cannot be moved to different locations. If buyers are not relocated, they can use money and capital to purchase goods from sellers. This relocation shock follows Townsend [1987]. Sellers are never relocated. During the final CM of their lives, buyers return to their original location and have access to all their remaining assets. Both buyers and sellers receive concave utility from consuming during the final CM of their lives. Monetary policy is implemented either by paying transfers to / raising taxes from the young or the old agents. The relocation shock creates a tradeoff between money and capital. Since capital pays a higher return than money unless the central bank runs the Friedman rule, capital is better suited as a store of value. However, because buyers can use money in the DM even if they are relocated, money is more liquid than capital - and the probability of relocation is a measure of the liquidity of capital, with capital being more liquid if relocation is less likely. As mentioned above, this setup blends standard OLG and LW frameworks - as in standard LW models, buyers and sellers trade with each other during the DM, and as in standard OLG models, young and old agents trade with each other during the CM. Having these two kinds of trades allows us to separate two properties of assets: Liquidity and store of value. In traditional OLG models with relocation, these two cannot easily be separated.

We first study a benchmark case where all buyers are relocated, meaning that capital is perfectly illiquid. In this version of the model, buyers face no tradeoff between money and capital, as capital can never be used to provide DM consumption, but since it (weakly) dominates in terms of rate of return, it is more useful to provide CM consumption. We show that in this case, CM consumption levels are independent of monetary policy and at the first-best level, and running the Friedman rule allows to implement the first-best consumption levels in the DM, but keeps the level of capital accumulation strictly below first best. Further, the level of capital accumulation is inversely related to aggregate hours worked in the CM, so hours worked in the CM are strictly above first best at the Friedman rule.

In the model with full relocation, there are two channels through which inflation affects capital accumulation: on the one hand, capital accumulation increases in inflation, because sellers are aware that they can sell less goods in the DM at higher inflation rates, so they accumulate more capital to provide for their CM consumption (*seller channel*); on the other hand, capital accumulation decreases in inflation since buyers hold less capital for CM consumption and tax payments if the real tax payment gets lower (*transfer channel*). The transfer channel is only active when old agents are taxed, so there is always a Mundell-Tobin effect when monetary policy is implemented by taxing the young agents, and it turns out that a constant money stock is optimal in this case. For any deflationary policy, the welfare loss from the reduction in capital accumulation (and therefore the increase in hours worked) is larger than the gains from increasing the consumption levels in the DM. If old agents are taxed instead, whether or not there is a Mundell-Tobin effect depends on the relative strength of the two channels, which is determined by the agents' preferences. If the elasticity of DM consumption is above one, there is a Mundell-Tobin effect, while if it is below one, there is actually a reverse Mundell-Tobin effect, meaning that capital accumulation decreases with inflation. Conversely, inflation reduces (increases) total hours worked in the CM if the elasticity is above (below) one. From this, it follows that the Friedman rule is optimal if the old are taxed and the elasticity of DM consumption is below one; if the elasticity is above one, the optimal money growth rate lies somewhere between the Friedman rule and one, and it is an increasing function of the elasticity of DM consumption in that interval. We also show that for any deflationary policy, welfare is higher if monetary policy is implemented over old buyers only. The reason is that under a deflationary policy, if buyers are taxed when old they can use the return on capital to pay parts of the tax.

Next, we analyze the full model with partially liquid capital, which means buyers face a tradeoff between money and capital regarding DM consumption. With money, buyers can always trade in the DM, but they need to forego some return if money growth is above the Friedman rule. By running the Friedman rule, the monetary authority is able to perfectly insure agents against the relocation shock, but then all DM trades are made with money, even though capital would be accepted in some of them. This adds a third channel through which inflation affects capital accumulation, which we call the *liquidity channel*. The higher the liquidity of capital, the more willingly buyers switch to accumulating capital instead of real balances if the return on money decreases. Because the liquidity channel strengthens the Mundell-Tobin effect, the Mundell-Tobin effect is more likely to occur at the Friedman rule for lower π, and in turn this makes it less likely that the Friedman rule is the optimal monetary policy, even if the old are taxed. In the limit where $\pi \to 0$, the Friedman rule is never optimal.

From the policymaker's point of view, the fundamental tradeoff is that the Friedman rule delivers efficiency in the DM, but a constant money stock is optimal regarding the CM. The reason for the latter point is that there are two ways to provide for (old-age) CM consumption: accumulation of capital when young, or transfers from young to old agents. The inherent return of such transfers is one, while the return of capital is larger than one; thus, the planner prefers capital accumulation whenever possible. If capital is not fully liquid, some DM trades need to be made with money, which implies that some of the sellers' CM consumption has to be financed with money. However, purchasing CM goods with money implies intergenerational transfers, as only young agents want to acquire money. Thus, setting the return on money equal to one, which can be achieved with a constant money stock, reflects the social return of using money (and thus intergenerational transfers) to acquire CM consumption. Running a constant money stock increases the price of DM consumption relative to the Friedman rule, and this increase in DM prices correctly reflects the externality of financing some of the sellers' consumption through intergenerational transfers - but this increase in the DM price also inefficiently lowers DM consumption. Thus, there is no single money growth rate that allows for (constrained) efficiency in both markets, and instead the optimal money growth rate depends on the probability that agents need to use money in the DM, and how they value DM consumption relative to labor disutility.

Existing literature. Aruoba and Wright [2003] is one of the earliest papers that includes capital in a LW framework. There, capital accumulation is independent of monetary policy if capital is fully illiquid. In our model, this is not true: because the OLG framework allows us to drop quasilinear preferences, capital accumulation is affected by monetary policy even in the version of our model with fully illiquid capital. Lagos and Rocheteau [2008] show that the Mundell-Tobin effect exists in LW models when capital is liquid. However, the Friedman rule still delivers the first-best outcome in their model, i.e., it simultaneously delivers optimal capital investment and efficient allocations in trades that require liquid assets. In Aruoba et al. [2011], capital reduces the cost of sellers to produce the DM good. They show that capital accumulation is affected more strongly by inflation if there is price-taking in the DM. Andolfatto et al. [2016] show that if taxes cannot be enforced and therefore the Friedman rule is not feasible, the first-best allocation can be implemented with a cleverly designed mechanism even if the capital stock is too small. In Wright et al. [2018, 2019], the authors study models where capital is traded in frictional markets, and they show that if money is needed to purchase capital, a reverse Mundell-Tobin effect can occur. In our model, a reverse Mundell-Tobin effect can also occur for some parameters, but due to preferences, not frictional markets. Gomis-Porqueras et al. [2020] show that there is a hump-shaped relationship between inflation and aggregate capital, as inflation affects capital accumulation negatively on the extensive margin by reducing the number of firms, besides the usual positive effect on the intensive margin.

There have been a few papers that find deviations from the Friedman rule to be optimal due to the Mundell-Tobin effect - e.g. Venkateswaran and Wright [2013], Geromichalos and Herrenbrueck [2017], Wright et al. [2018], or Altermatt [2019a]. However, in these papers there is usually an additional friction that leads to underinvestment at the Friedman rule, e.g., limited pledgeability, taxes, or wage bargaining. If these frictions are shut down in the papers mentioned, the Mundell-Tobin effect still exists, but the Friedman rule is optimal. In this paper, the Friedman rule is optimal if and only if there is no Mundell-Tobin effect.

In the OLG literature following Wallace [1980], the Mundell-Tobin effect has also been studied. Azariadis and Smith [1996] show that if there is private information about an agent's type, a Mundell-Tobin effect exists for low levels of inflation, while a reverse Mundell-Tobin effect exists for high levels of inflation. In their model, agents are either borrowers or lenders, and higher inflation makes bank deposits a relatively less attractive means of saving, which increases a savers' value of misrepresenting his type and defaulting on a loan. To prevent this, banks ration loans, which depresses the borrowers' ability to accumulate capital. In models with relocation shocks, Smith [2002, 2003] and Schreft and Smith [2002] have claimed to show that the Friedman rule is suboptimal because of the Mundell-Tobin effect.[4] However, OLG models typically find deviations from the Friedman rule to be optimal even without the Mundell-Tobin effect[5], as in Weiss [1980], Abel [1987], or Freeman [1993]. Bhattacharya et al. [2005] and Haslag and Martin [2007] build on these results to show that the results in Smith [2002] and the other papers mentioned are not driven by the Mundell-Tobin effect, but by the standard properties of the OLG models. The debate whether the Mundell-Tobin effect itself can render deviations from the Friedman rule to be optimal in an OLG environment thus remained unsettled.

Zhu [2008] was the first to combine the OLG and LW structures. In his model, agents do not know their type during the first CM when they are able to accumulate assets. Therefore, the Friedman rule can be suboptimal for some parameters, as it makes saving relatively cheap and

[4]See also Schreft and Smith [1997], which focuses on positive inflation rates, but endogenizes the return on capital.

[5]There is a further complication in the welfare analysis of OLG models due to the absence of a representative agent. Freeman [1993] shows that the Friedman rule is typically Pareto optimal, but not maximizing steady state utility in OLG models. In this paper, we are going to focus on steady-state optimality when analyzing optimal policies in OLG models.

therefore reduces the sellers' willingness to produce in the DM. In contrast to this, our model follows Altermatt [2019b] by assuming that each agent knows his type. Hiraguchi [2017] extends the model of Zhu [2008] by including capital and shows that the Friedman rule remains suboptimal in this case. In another recent paper that combines OLG and LW, Huber and Kim [2020] show that the Friedman rule can be suboptimal if old agents face a higher disutility of labor than young agents.

We think that our paper contributes to the existing literature in a number of important ways. First, it is able to reconcile Smith [2002] with Bhattacharya et al. [2005] and Haslag and Martin [2007], by showing that even in a model with an OLG structure, the Friedman rule can be optimal for some parameters, but that it is never optimal when there is a Mundell-Tobin effect. Second, our paper shows that in a New Monetarist model without quasilinear preferences, capital accumulation is affected by monetary policy even if capital is illiquid, and capital accumulation can be inefficiently low at the Friedman rule due to the Mundell-Tobin effect. Third, we made some advances in understanding the frictions that arise in a model that combines OLG and LW, most importantly by showing that the timing of monetary policy implementation matters for welfare in these kind of models.

Outline. The rest of this paper is organized as follows. In Section 4.1, the environment and the planner's solution is explained. In Section 3.2, we present the market outcome for perfectly liquid capital, and in Section 3.3, we discuss the market outcome for perfectly illiquid capital and monetary policy implementation. Section 3.4 presents the results of the full model, and finally, Section 3.5 concludes.

3.1 The model

Our model is a combination of the environment in Lagos and Wright [2005], and the overlapping generations model (OLG) with relocation shocks from Townsend [1987], as used by Smith [2002]. Time is discrete and continues forever. Each period is divided into two subperiods, called the decentralized market (DM) and the centralized market (CM). There are two distinct locations, which we will sometimes call islands. The two locations are completely symmetric, and everything we describe happens simultaneously on both islands. At the beginning of a period, the CM takes place, and after it closes, the DM opens and remains open until the period ends. At the beginning of each period, a new generation of agents is born, consisting of one unit mass per island each of buyers and sellers. An agent born in period t lives until the end of the CM in period $t+1$. Each generation is named after the period it is born in. Figure 3.1 gives an overview of the sequence of

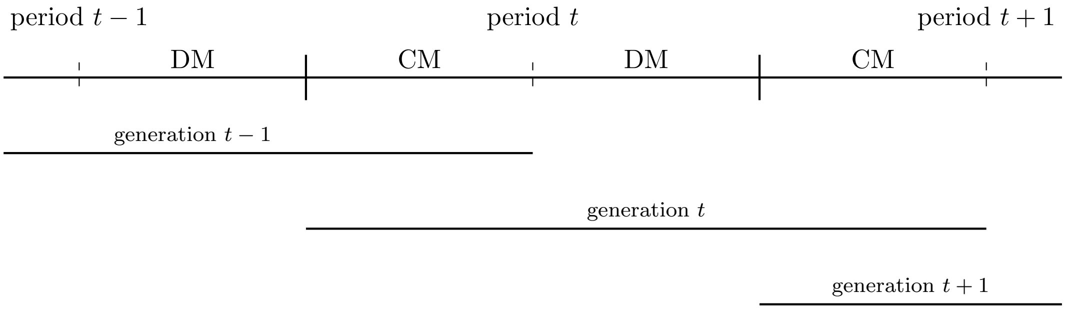


Figure 3.1: Timeline with lifespans of generations.

subperiods and the lifespans of generations. There is also a a monetary authority.

Both buyers and sellers are able to produce a general good x during the first CM of their life at linear disutility h, whereas incurring the disutility h yields exactly h units of general goods; buyers and sellers also both receive utility from consuming the general good during the second CM of their life. During the DM, sellers are able to produce special goods q at linear disutility; buyers receive utility from consuming these special goods.

The preferences of buyers are given by

$$\mathbb{E}_t\{-h_t^b + u(q_t^b) + \beta U(x_{t+1}^b)\}. \tag{3.1}$$

Equation (3.1) states that buyers discount the second period of their life by a factor $\beta \in (0,1)$, gain utility $u(q)$ from consuming the special good in the DM and $U(x)$ from consuming the general good in the CM, with $u(0) = 0$, $u'(q) > 0$, $u''(q) < 0$, $u'(0) = \infty$, $U(0) = 0$, $U'(x) > 0$, $U''(x) < 0$, $U'(0) = \infty$, and linear disutility h from producing the general good during their first CM.[6] The preferences of the sellers are

$$-h_t^s - q_t^s + \beta U(x_{t+1}^s). \tag{3.2}$$

Sellers also discount the second period of their life by a factor β, gain utility $U(x)$ from consuming in the CM and disutility q from producing in the DM, with $\bar{q} = u(\bar{q})$ for some $\bar{q} > 0$.

During the CM, general goods can be sold or purchased in a centralized market. During the DM, special goods are sold in a centralized market. A fraction π of buyers are relocated during the

[6]We also assume strictly convex marginal utility in the CM, i.e. $U'''(x) > 0$. Most commonly used utility functions satisfy this assumption and it simplifies the proofs for the last section.

DM, meaning that they are transferred to the other island without the ability to communicate with their previous location. Sellers are not relocated, and during the CM, no relocation occurs for both types of agents. Relocated buyers are transferred back to their original location for the final CM of their life.[7] Relocation occurs randomly, so for an individual agent, the probability of being relocated is π. Buyers learn at the beginning of the DM whether they are relocated or not.

The monetary authority issues fiat money M_t, which it can produce without cost. The monetary authority always implements its policies at the beginning of the CM. The amount of general goods that one unit of fiat money can buy in the CM of period t is denoted by ϕ_t. The gross inflation rate is defined as ϕ_t/ϕ_{t+1}, and the growth rate of fiat money from period $t-1$ to t is $\frac{M_t}{M_{t-1}} = \gamma_t$. Monetary policy is implemented by issuing newly printed fiat money either to young or to old buyers via lump-sum transfers (or lump-sum taxes in the case of a decreasing money stock).[8] We denote transfers to young buyers as τ^y, and transfers to old buyers as τ^o. Furthermore, we will use an indicator variable $\mathcal{I}$ to denote the regime, i.e., which generation is taxed. If $\mathcal{I} = 1$ ($\mathcal{I} = 0$), young buyers (old buyers) are taxed, which means τ^y (τ^o) is set such that the money growth rate γ_t chosen by the monetary authority can be implemented, while $\tau^o = 0$ ($\tau^y = 0$).

Besides fiat money, there also exists capital k in this economy. During the CM, agents can transform general goods into capital. One unit of capital delivers $R > 1$ units of real goods in the CM of the following period. Capital is immobile, meaning that it is impossible to move capital to other locations during the DM. It is also impossible to create claims on capital that can be verified by other agents. We will assume throughout the paper that

$$R\beta = 1. \tag{3.3}$$

As we will see in the planner's problem below, this assumption implies that accumulating capital is more efficient than intergenerational transfers at financing old-age consumption.[9] It also implies

[7]In Smith [2002], each agent lives only for two periods. Relocation occurs during the last period of an agent's life, meaning that all assets that he cannot spend during that period are wasted from his point of view. Our model crucially differs from Smith [2002] in that regard, as our agents have access to all their assets during the final period of their life.

[8]As we will show in this paper, the exact timing of the lump-sum taxes is irrelevant for consumption allocations, but not for welfare. Assuming that only buyers are taxed is without loss of generality.

[9]For all our results to go through, $R > 1$ actually suffices. However, assuming $R\beta = 1$ has the added benefit that two common definitions of the Friedman rule coincide, i.e., the Friedman rule is given by $\gamma = \beta = \frac{1}{R}$. In LW models, the Friedman rule is typically defined as setting the money growth rate equal to the discount factor ($\gamma = \beta$), while in OLG models, the Friedman rule is typically defined as setting the rate of return on money equal to the return on other assets in the economy ($\frac{1}{\gamma} = R$). We think that the second definition is the right one in the context of our model, as it fits most closely the original definition of setting the opportunity cost of holding money to zero, and

that capital is a good investment, which introduces a tradeoff between liquid money and non-(or partially)-liquid capital in terms of return.

A Mundell-Tobin effect means that capital investments are increasing with inflation. Thus we speak of a Mundell-Tobin effect in the model if $\frac{\partial K}{\partial \gamma} > 0$, where $K = k^b + k^s$ are the total capital investments of buyers and sellers. Conversely, we speak of a reverse Mundell-Tobin Effect if $\frac{\partial K}{\partial \gamma} < 0$ i.e., if total capital decreases in inflation.

3.1.1 Planner's problem

For the planner's problem, we focus on maximizing steady-state welfare of a representative generation, while ignoring the initial old. By doing so, we follow papers like Smith [2002] and Haslag and Martin [2007], as we want to compare our results to theirs.

The planner maximizes the utility of a representative generation, which is given by

$$V^g = -h^b - h^s - q^s + \pi\left(u(q^m) + \beta U(x^m)\right) + (1-\pi)\left(u(q^b) + \beta U(x^b)\right) + \beta U(x^s)), \tag{3.4}$$

where a superscript m denotes consumption of relocated buyers (movers). DM-consumption of buyers must be financed by transfers from sellers, so $\pi q^m + (1-\pi)q^b = q^s$. To finance CM-consumption, the planner has two possibilities. Either the young agents work for the old and CM consumption is financed by transfers, or young agents work in order to invest in capital, and consume the returns when old. The implied return of an intergenerational transfer is 1, as when goods are taken from young agents and given to old agents, the additional goods produced by a young agent of a representative generation equal the additional goods consumed by an old agent of a representative generation. Since $R > 1$ from (3.3), it requires strictly less work to provide the same amount of CM consumption through capital investment instead of intergenerational transfers, so a planner wants to finance all CM consumption through capital investment. Taking this into account and using $H = h^b + h^s$ to denote total labor supply, the planner's problem is

$$\max_{H,K,q^b,q^m,q^s,x^b,x^m,x^s} -H - q^s + \pi\left(u(q^m) + \beta U(x^m)\right) + (1-\pi)\left(u(q^b) + \beta U(x^b)\right) + \beta U(x^s))$$

$$s.t. \qquad \pi q^m + (1-\pi)q^b = q^s$$

since $\frac{1}{R}$ shows up in our relevant first-order conditions, not β. But to make it clear that our results don't depend on a debatable definition of the Friedman rule, we assume $R\beta = 1$ throughout the paper.

$$H = K$$

$$\pi x^m + (1-\pi)x^b + x^s = RK.$$

Thus the first-best levels of DM and CM consumption q^* and x^*, hours worked H^* and capital investment K^* solve:

$$q^b = q^m = q^s = q^* \quad \text{solving} \quad u'(q^*) = 1 \tag{3.5}$$

$$x^b = x^m = x^s = x^* \quad \text{solving} \quad U'(x^*) = \frac{1}{\beta R} \tag{3.6}$$

$$H^* = K^* = \frac{2x^*}{R}. \tag{3.7}$$

We will later use these results as a benchmark to compare market outcomes against.[10]

3.1.2 Market outcomes

In the DM, special goods are sold in competitive manner.[11] Due to anonymity and a lack of commitment, all trades have to be settled immediately. Therefore, buyers have to transfer assets to sellers in order to purchase special goods. Because capital cannot be transported to other locations and claims on capital are not verifiable, relocated buyers can only use fiat money to settle trades. Nonrelocated buyers can use both fiat money and capital to purchase special goods. We will use p_t to denote the price of special goods in terms of fiat money. All buyers face the same price, regardless of their means of payment. As sellers are not relocated during the DM, all of them accept both fiat money and capital of nonrelocated buyers as payment. Because the problem is symmetric, we will only focus on one location for the remainder of the analysis.

Buyer's lifetime problem

A buyer's value function at the beginning of his life is given by

$$V^b = \max_{h_t, q_t^m, q_t^b, x_{t+1}^m, x_{t+1}^b} \quad -h_t + \pi\big(u(q_t^m) + \beta U(x_{t+1}^m)\big) + (1-\pi)\big(u(q_t^b) + \beta U(x_{t+1}^b)\big)$$

[10]With intergenerational transfers, the optimal CM-consumption levels are $x^b = x^m = x^s = x_1$ solving $U'(x_1) = \frac{1}{\beta}$, with total labor supply given by $2x_1$. However, if the same amount of CM consumption is financed by capital total labor supply would be only $\frac{2x_1}{R}$. Thus for $R > 1$ it is more efficient to finance CM consumption with capital.

[11]Zhu [2008] studies an economy with bilateral meetings and ex-ante uncertainty about an agent's type in a model that is otherwise similar to ours, and shows that these frictions can make devations from the Friedman rule optimal under some conditions. By assuming fixed types and competitive markets, we want to highlight that our results stem from different frictions than those found by Zhu.

$$
\begin{aligned}
s.t.\quad h_t + \mathcal{I}\tau_t^y &= \phi m_t + k_t^b \\
p_t q^m &\le m_t \\
p_t q^b &\le m_t + \frac{Rk^b}{\phi_{t+1}} \\
x_{t+1}^m &= \phi_{t+1} m_t + Rk_t^b - \phi_{t+1} p_t q_t^m + (1-\mathcal{I})\tau_t^o \\
x_{t+1}^b &= \phi_{t+1} m_t + Rk_t^b - \phi_{t+1} p_t q_t^b + (1-\mathcal{I})\tau_t^o.
\end{aligned}
$$

All variables with a superscript m indicate decisions of relocated buyers (movers). Variables with superscript b indicate decisions of buyers prior to learning about relocation, or those of buyers that aren't relocated, depending on the context. The first constraint is the standard budget constraint for the portfolio choice when young. The second constraint denotes that relocated buyers cannot spend more than their money holdings during the DM, and the third constraint denotes that nonrelocated buyers cannot spend more than their total wealth for consumption during the DM.[12] The fourth and fifth constraint denote that buyers use all remaining resources for consumption when old.

We can simplify the problem by substituting some variables. Additionally, we only consider inflation rates where capital (weakly) dominates money in terms of return, i.e. $\frac{\phi_t}{\phi_{t+1}} \ge \frac{1}{R}$. In this case, the second constraint always holds at equality, as there is no reason for buyers to save money for the CM if capital pays a higher return. We also know that the third constraint never binds, because spending all wealth during the DM would imply $x_{t+1} = 0$, but this violates the Inada conditions. After simplification, the buyer's problem is

$$
\begin{aligned}
V^b = \max_{m_t, k_t^b, q_t^b} \quad & \mathcal{I}\tau^y - \phi_t m_t - k_t^b + \pi \left(u\left(\frac{m_t}{p_t}\right) + \beta U(Rk_t^b + (1-\mathcal{I})\tau^o) \right) \\
& + (1-\pi)\big(u(q_t^b) + \beta U(\phi_{t+1} m_t + Rk_t^b - \phi_{t+1} p_t q_t^b) + (1-\mathcal{I})\tau^o)\big).
\end{aligned} \tag{3.8}
$$

Seller's lifetime problem

A seller's value function at the beginning of his life is given by

$$
V^s = \max_{h_t, q_t^s, x_{t+1}^s} \quad - h_t^s - q_t^s + \beta U(x_{t+1}^s)
$$

[12] The purchasing power of capital is scaled by $\frac{R}{\phi_{t+1}}$ to ensure that buyers give up the same amount of CM consumption by paying with capital and money.

$$s.t. \quad h_t^s = k_t^s$$
$$x_{t+1}^s = Rk_t^s + \phi_{t+1} p_t q_t^s.$$

Here, we already assumed that sellers do not accumulate money in the first CM, which is true in equilibrium for $\phi_t/\phi_{t+1} \geq \frac{1}{R}$. Thus, the first constraint denotes that sellers work only to accumulate capital, and the second constraint denotes that a seller's CM consumption is equal to the return on capital plus his revenue from selling the special good in the DM. Again, we can simplify the problem by substituting in the constraints. After simplification, the seller's problem is

$$V^s = \max_{q_t^s, k_t^s} \quad -k_t^s - q_t^s + \beta U(Rk_t^s + \phi_{t+1} p_t q_t^s). \tag{3.9}$$

3.2 Equilibrium with perfectly liquid capital

In this section, we solve for the market equilibrium in the special case of $\pi = 0$, which means that no relocation occurs during the DM. This case represents perfectly liquid capital, as all buyers can use capital during the DM, and abstracts from any uncertainty for all agents in the model. As money and capital are equally liquid and safe in this case, only the rate of return of the assets matters, and agents will only hold the asset with the higher rate of return. For $\phi_t/\phi_{t+1} \geq \frac{1}{R}$, capital is the asset with the (weakly) higher rate of return, and as this is the case we are most interested in, we abstract from money (and monetary policy) in this section. As we used p_t to denote the price of the DM good in terms of fiat money, we have to alter the problem slightly, as this price is undefined if money is not held in equilibrium. In this section, we therefore introduce ρ_t, which is the price of the DM good in terms of capital.

Given these alterations of the model, the buyer's problem from equation (3.8) becomes

$$V^b = \max_{k_t^b, q_t^b} \quad -k_t^b + u(q_t^b) + \beta U((k_t^b - \rho_t q_t^b)R),$$

and yields the following first-order conditions:

$$q^b: \quad u'(q^b) = \rho_t \beta R U'((k_t^b - \rho_t q_t^b)R) \tag{3.10}$$

$$k^b: \qquad 1 = \beta R U'((k_t^b - \rho_t q_t^b)R). \tag{3.11}$$

The seller's problem is only affected by the change in notation. Solving equation (3.9) yields

$$q^s: \quad 1 = \beta R \rho_t U'((k_t^s + \rho_t q_t^s)R) \tag{3.12}$$

$$k^s: \quad 1 = \beta R U'((k_t^s + \rho_t q_t^s)R). \tag{3.13}$$

We already see from equations (3.11) and (3.13) that CM consumption is equal to the first-best level in this equilibrium i.e. $x^b = x^s = x^*$. Next we show that also DM consumption is at the first best level, i.e. $q^b = q^s = q^*$. Combining equations (3.12) and (3.13) gives $\rho_t = 1$ assuming optimal capital holdings of sellers are interior,[13] which means that DM prices are such that the seller is indifferent between working in the CM or the DM. Then, combining this with equations (3.10) and (3.11) yields

$$u'(q^b) = 1.$$

Furthermore, it is easily confirmed that labor supply and total capital investments are also at their first-best levels: $h^s + h^b = k^s + k^b = \frac{2x^*}{R} = H^* = K^*$. Thus, we can conclude that perfectly liquid capital allows to implement the planner's solution.

3.3 Equilibrium with perfectly illiquid capital

Having shown that there are no inefficiencies with perfectly liquid capital, we now want to investigate the other extreme case, which is perfectly illiquid capital. In the model, this is captured by $\pi = 1$, which means that all buyers are relocated during the DM. In this case, fiat money is the only way to acquire consumption during the DM. Thus, for $\phi_t/\phi_{t+1} \geq \frac{1}{R}$, buyers face no tradeoff between holding fiat money and capital, as only fiat money allows them to acquire DM consumption, while capital (weakly) dominates in terms of providing CM consumption.

[13]To implement the planner's solution with perfectly liquid capital, utility functions have to be such that sellers want to consume at least as much in the CM as they receive from selling the efficient amount of special goods at $\rho = 1$ in the DM while holding no capital. Thus we need $x_s = x^* > q^* R$ or $U'(q^* R) \geq \frac{1}{\beta R}$. We are assuming that this holds for the remainder of the paper. An alternative assumption we could make to prevent this issue is that the measure of sellers is sufficiently larger than the measure of buyers, such that individual sellers don't sell too many special goods in the DM. This friction might be interesting to study in other contexts, but it is not relevant for the points we want to make in this paper.

With $\pi = 1$, the buyer's lifetime value function (3.8) simplifies to

$$V^b = \max_{m_t, k_t^b} \quad \mathcal{I}\tau^y - \phi_t m_t - k_t^b + u\left(\frac{m_t}{p_t}\right) + \beta U(Rk_t^b + (1-\mathcal{I})\tau^o).$$

Solving this problem yields two first-order conditions:

$$m_t : \qquad p_t\phi_t = u'\left(\frac{m_t}{p_t}\right) \tag{3.14}$$

$$k_t : \qquad \frac{1}{\beta R} = U'(Rk_t^b + (1-\mathcal{I})\tau^o), \tag{3.15}$$

while solving the seller's problem yields the following first-order conditions:

$$q^s : \quad 1 = \phi_{t+1} p_t \beta_t U'(Rk_t^s + p_t q_t^s \phi_{t+1}) \tag{3.16}$$

$$k^b : \quad 1 = \beta R U'(Rk_t^s + p_t q_t^s \phi_{t+1}). \tag{3.17}$$

Combining equations (3.16) and (3.17) yields $p_t = \frac{R}{\phi_{t+1}}$. Plugging this into equation (3.14) gives

$$u'(q_t^m) = \frac{\phi_t}{\phi_{t+1}} R. \tag{3.18}$$

Equations (3.15) and (3.17) demonstrate that the CM consumption is always at the first-best level, independent of monetary policy $x^b = x^s = x^*$.

Given the first-order conditions, we first derive the stationary equilibrium when monetary policy is implemented over young buyers ($\mathcal{I} = 1$). In a stationary equilibrium we must have: $q^m = q^s$ (DM market clearing), $m = M$ (money market clearing) and $\phi/\phi_{+1} = \gamma$ i.e. the inflation rate must equal the growth rate of the money supply since the real value of money is constant over time, implying $\phi M = \phi_{+1} M_{+1}$. Furthermore the real value of taxes/transfers paid/received by young buyers is given by: $\tau^y = \phi(M - M_{-1}) = \frac{\gamma - 1}{\gamma}\phi M$. Using this and the definitions and first-order conditions derived above for $\pi = 1$, we can then define a stationary equilibrium with perfectly illiquid capital as a list of eight variables $\{h^b, h^s, k^b, k^s, \phi_{+1}M, q^m, x^b, x^s\}$ solving:

$$u'(q^m) = \gamma R \tag{3.19}$$

$$x^b = x^s = x^* \tag{3.20}$$

$$\phi_{+1}M = q^m R \tag{3.21}$$
$$k^{b,\mathcal{I}=1} = \frac{x^*}{R} \tag{3.22}$$
$$h^{b,\mathcal{I}=1} = q^m R + \frac{x^*}{R} \tag{3.23}$$
$$k^s = \frac{x^*}{R} - q^m \tag{3.24}$$
$$h^s = k^s. \tag{3.25}$$

Equation (3.20) shows that x^b is independent of the inflation rate and thus inflation does not affect the buyer's capital accumulation. However, equation (3.24) shows that total capital accumulation is still indirectly affected by inflation. Sellers accumulate less capital if they expect to sell more goods in the DM - and DM consumption and thus also real balances are decreasing in the inflation rate, as we know from equation (3.19). Differentiating both sides of (3.19) with respect to γ yields:

$$\frac{\partial q^m}{\partial \gamma} = \frac{R}{u''(q^m)} < 0, \tag{3.26}$$

which is negative from the strict concavity of $u(q)$. In turn, this implies from equation (3.24) that seller's capital accumulation is increasing in inflation. This is the first channel through which capital accumulation is affected by the inflation rate, and it is active independent of the tax regime. Since it affects the sellers' capital accumulation, we call it the *seller channel*.

Total capital investment and labor supply are given by:

$$K^{\mathcal{I}=1} = k^{b,\mathcal{I}=1} + k^s = \frac{2x^*}{R} - q^m \tag{3.27}$$
$$H^{\mathcal{I}=1} = h^{b,\mathcal{I}=1} + h^s = \frac{2x^*}{R} + q^m(R-1) = K^{\mathcal{I}=1} + q^m R. \tag{3.28}$$

Next, we derive the stationary equilibrium when monetary policy is implemented over old buyers ($\mathcal{I}=0$). The real value of taxes/transfers paid/received by old buyers is given by: $\tau^o = \phi_{+1}(M_{+1} - M) = \phi_{+1}M(\gamma-1)$. Thus under stationarity $\tau^y = \tau^o = \tau$ for a given γ. Equilibrium consumption levels are unaffected by the different monetary policy regimes. The only changes in the equilibrium allocation affect the labor supply (equation (3.23)) and capital accumulation of buyers (equation (3.22)). These now read:

$$h^{b,\mathcal{I}=0} = \gamma q^m R + \frac{x^*}{R} - q^m(\gamma-1) \tag{3.29}$$
$$k^{b,\mathcal{I}=0} = \frac{x^*}{R} - q^m(\gamma-1), \tag{3.30}$$

and aggregate capital and labor supply:

$$K^{\mathcal{I}=0} = \frac{2x^*}{R} - \gamma q^m \tag{3.31}$$

$$H^{\mathcal{I}=0} = \frac{2x^*}{R} - \gamma q^m + \gamma R q^m = K^{\mathcal{I}=0} + \gamma R q^m. \tag{3.32}$$

Compare this with the equations (3.23) and (3.22) resp. (3.27) and (3.28) for the model with monetary policy implemented over young buyers. If there is inflation ($\gamma > 1$), $K^{\mathcal{I}=0} < K^{\mathcal{I}=1}$ and $H^{\mathcal{I}=0} > H^{\mathcal{I}=1}$, i.e. capital accumulation is lower when the old buyers are receiving the transfers, while the hours worked when young are higher. If there is deflation ($\gamma < 1$), $K^{\mathcal{I}=0} > K^{\mathcal{I}=1}$ and $H^{\mathcal{I}=0} < H^{\mathcal{I}=1}$, i.e., capital accumulation is higher when the old pay the taxes (a deflationary policy means $\tau < 0$), while the hours worked when young are lower. For $\gamma = 1$, the two equilibria coincide. To put this in other words, if monetary policy consists of making transfers to buyers ($\gamma > 1$), total work can be kept lower if monetary policy is implemented over young buyers, whereas the opposite is true if the monetary authority wants to implement a deflationary policy and thus raises taxes. The reason for this is that capital has a return $R > 1$; thus, if agents receive a transfer, it is better to receive it when young and invest it in capital, whereas if agents have to pay a tax, it is better to use the return on capital to pay it when old instead of paying it directly from labor income when young.

This also shows that with $\mathcal{I} = 0$, there is a second channel through which inflation affects capital accumulation: Equation (3.30) shows that the buyers' capital accumulation is a function of q^m, which is decreasing in inflation, and of the inflation rate itself. Since from equation (3.22) we know that capital accumulation of buyers is independent of inflation for $\mathcal{I} = 1$, we can conclude that this channel arises due to the monetary policy regime. We call this the *transfer channel.*

Before we characterize welfare for general inflation rates, we want to analyze what happens at the Friedman rule.

Proposition 3.1. *For $\pi = 1$, both DM and CM consumption are at the first best level at the Friedman rule ($\gamma = 1/R$), i.e. $q^m = q^*$ and $x^s = x^b = x^*$. Total hours worked are strictly above the first-best level, and strictly higher if the Friedman rule is implemented over taxes on the young buyers, compared to implementation through taxes on the old buyers; i.e., $H^{\mathcal{I}=1}|_{FR} > H^{\mathcal{I}=0}|_{FR} > H^*$.*

It can easily be seen from equation (3.19) that DM consumption is at the first-best level for $\gamma = 1/R$, and equation (3.20) shows that CM consumption is always at the first best level. Thus, consumption is efficient at the Friedman rule. Total capital and total hours worked however are not efficient. From (3.27) we see that $K^{\mathcal{I}=1}|_{FR} = 2x^*/R - q^*$ and $K^{\mathcal{I}=0}|_{FR} = 2x^*/R - q^*/R$ from (3.31). Thus capital investment is too low in both monetary policy regimes compared to the first-best level. This is not surprising, as sellers can only be compensated with money, which implies that their CM consumption will be partially financed through transfers from young to old agents - old sellers enter the CM with money and use it to purchase consumption goods from young buyers. At the first best, all CM consumption is financed with capital investment, so these intergenerational transfers imply an inefficiency. The inefficiency shows up in the aggregate labor supply, which is too high compared to the first best:

$$H^{\mathcal{I}=1}|_{FR} = 2x^*/R + q^*(R-1) > H^{\mathcal{I}=0}|_{FR} = 2x^*/R + q^*(R-1)/R > H^*.$$

Since $R > 1$, it can easily be seen that implementing the Friedman rule by taxing old buyers is more efficient - it allows to achieve the same consumption levels at strictly lower hours worked.[14] Proposition 3.1 shows that even though consumption is at the first-best level at the Friedman rule, there is still a welfare loss from hours worked, so it is not obvious that running the Friedman rule is welfare-maximizing - and as we will show, it turns out that the Friedman rule is only the welfare-maximizing policy under some conditions in this economy.

Next, we investigate the effects of inflation on total labor supply $H = h^b + h^s$ and total capital accumulation $K = k^b + k^s$ for both policy implementation schemes. As we will show, these effects depend on the absolute value of the elasticity of DM consumption, which we denote as $|\varepsilon_{q^m}|$, and on the coefficient of relative risk aversion, which we denote as $\eta(q) = -\frac{qu''(q)}{u'(q)}$.

Proposition 3.2. *With $\mathcal{I} = 0$ inflation affects total labor supply and total capital accumulation in the following way:*

1. *If $|\varepsilon_{q^m}| < 1$: $\frac{\partial H^{\mathcal{I}=0}}{\partial \gamma} > 0$ and $\frac{\partial K^{\mathcal{I}=0}}{\partial \gamma} < 0$: Reverse Mundell-Tobin effect.*
2. *If $|\varepsilon_{q^m}| > 1$: $\frac{\partial H^{\mathcal{I}=0}}{\partial \gamma} < 0$ and $\frac{\partial K^{\mathcal{I}=0}}{\partial \gamma} > 0$: Mundell-Tobin effect.*
3. *If $|\varepsilon_{q^m}| = 1$: $\frac{\partial H^{\mathcal{I}=0}}{\partial \gamma} = 0$ and $\frac{\partial K^{\mathcal{I}=0}}{\partial \gamma} = 0$: No Mundell-Tobin effect.*

[14] We are assuming in this Friedman rule equilibrium that agents finance their CM consumption with capital, even though they are indifferent between money and capital on an individual level. For the economy as a whole, financing CM consumption with money would be much more inefficient. In the extreme case where all CM-consumption is financed with money, we have a pure monetary economy where agents don't invest into capital. In this case consumption in the DM and the CM would still be efficient at the Friedman rule. But total work would be excessive since buyers would have to provide the total real return on money of $1/\gamma = R$ by working.

where $|\varepsilon_{q^m}| = \frac{1}{\eta(q^m)}$. *With* $\mathcal{I} = 1$, $\frac{\partial H^{\mathcal{I}=0}}{\partial \gamma} < 0$ *and* $\frac{\partial K^{\mathcal{I}=0}}{\partial \gamma} > 0$ *and thus the Mundell-Tobin effect holds for all values of* $|\varepsilon_{q^m}|$.

The proof to this Proposition can be found in the appendix. As we have pointed out above, there are two channels through which capital accumulation is affected by the inflation rate when $\pi = 1$. Through the seller channel, inflation has a positive effect on capital accumulation. Sellers understand that buyers want to consume less DM goods at higher inflation rates, so they finance a larger share of their CM consumption through capital investment. Since their CM consumption is constant at x^*, financing a larger share of it through capital investment implies higher capital investment. The seller channel is independent of the monetary policy regime.

The transfer channel is only active for $\mathcal{I} = 0$. With fully illiquid capital, buyers accumulate capital only for CM consumption and expenditures when old. Since CM consumption is constant at x^*, the buyers' capital accumulation is independent of inflation if they don't pay taxes / receive transfers when old, which is the case when $\mathcal{I} = 1$. With $\mathcal{I} = 0$ however, the real value of the tax to be paid or transfers received when old varies with the inflation rate, and thereby affects the buyers' capital accumulation. For $\gamma < 1$ (deflation), buyers have to pay a tax when old, so they need to accumulate more capital in order to provide for x^* and the tax payment. If inflation is positive instead, buyers receive a transfer when old, so they can partly finance x^* through the transfer and need to accumulate less capital. The effect of inflation on capital accumulation through the transfer channel has two components: On the one hand, higher inflation (less deflation) increases the nominal value of the transfer (decreases the nominal value of the tax), but on the other hand, higher inflation (less deflation) decreases the value of money. For $\gamma < 1$, less deflation decreases the nominal value of the tax and also the real value of money. Thus the real tax payment decreases and buyers hold less capital. For positive inflation rates, either effect can dominate, depending on the elasticity of DM consumption.[15]

Proposition 3.2 describes the effect on aggregate capital accumulation, which is given by the net effect of the two channels. Since the transfer channel is shut down for $\mathcal{I} = 1$, capital accumulation is always increasing in inflation when taxes are paid by the young, so there is a Mundell-Tobin effect. For $\mathcal{I} = 0$, the aggregate effect depends on the elasticity of DM consumption $|\varepsilon_{q^m}|$, as this governs both the sign of the transfer channel at higher inflation rates, and the relative strength of the two channels. With $\pi = 1$, the elasticity of DM consumption is fully determined by the coefficient of relative risk aversion $\eta(q^m)$. If $|\varepsilon_{q^m}| = 1$, the increase in capital accumulation with inflation through the seller channel is exactly offset by a decrease in capital accumulation coming

[15]Specifically, if $|\varepsilon_{q^m}| > \frac{\gamma}{\gamma - 1}$, the effect through the transfer channel of inflation on capital accumulation is positive, so a positive correlation is more likely for high DM elasticity and higher inflation rates.

from the transfer channel, and the aggregate capital investment remains constant. Thus, there is no Mundell-Tobin effect in this case. With $|\varepsilon_{q^m}| > 1$, the effect through the seller channel is strong, while there is only a weak negative effect on capital accumulation from the transfer channel at low inflation rates, so there is a Mundell-Tobin effect on aggregate. The reason that this happens for high values of elasticity is that in this case, buyers reduce their DM consumption by a lot if inflation increases, so in turn the sellers' capital accumulation is reacting strongly to changes in inflation. Strong changes in DM consumption also imply that the real value of the tax/transfer decreases rapidly as inflation increases, which weakens the negative effect on capital accumulation from the transfer channel. The contrary is true for $|\varepsilon_{q^m}| < 1$: DM consumption changes very little as inflation varies, implying that sellers' capital accumulation also varies very little with inflation. On the other hand, higher inflation rates leave the value of money almost unchanged, so the transfer channel has a strong negative effect on capital accumulation. On aggregate, the negative effect from the transfer channel dominates, such that there is a reverse Mundell-Tobin effect for aggregate capital accumulation.

Proposition 3.2 also shows that a Mundell-Tobin effect is correlated with a negative effect on aggregate labor supply. This may seem counterintuitive, as from equations (3.28) and (3.32) capital seems to *increase* total labor in the CM. To understand the connection look at market clearing in the CM in the model with $\mathcal{I} = 1$ (without loss of generality). The total amount of goods consumed in the CM is $2x^*$. Out of these, $K^{\mathcal{I}=1}R$ are produced from capital investments in the previous period, and $\phi M + \tau = \phi_{+1}M$ come from intergenerational transfers - the amount which equals the money holdings of young buyers. To sum up:

$$2x^* = K^{\mathcal{I}=1}R + \phi_{+1}M. \tag{3.33}$$

Now suppose capital increases because inflation increases and there is a Mundell-Tobin effect. By (3.33), this decreases real money holdings, as $\phi_{+1}M = 2x^* - RK^{\mathcal{I}=1}$ and x^* is unaffected by inflation. By (3.28), the effect on total labor supply is then:

$$H^{\mathcal{I}=1} = K^{\mathcal{I}=1} + \phi_{+1}M = K^{\mathcal{I}=1} + 2x^* - K^{\mathcal{I}=1}R = 2x^* - K^{\mathcal{I}=1}(R-1).$$

Thus if capital increases, this *decreases* total labor because real money holdings decrease more than capital increases. The intuition is again that financing CM-consumption with capital instead of transfers is more efficient in this economy because $R > 1$, and therefore a shift in financing CM consumption from intergenerational transfers to capital accumulation decreases total labor - with more capital, the same amount of CM consumption can be achieved with less labor.

After having established the effects of inflation on capital and labor supply, we can derive the optimal monetary policy.

Proposition 3.3. *With $\pi = 1$, the optimal money growth rate under $\mathcal{I} = 0$ is $\gamma^* = 1/R$ for $|\varepsilon_{q^m}| \leq 1$, and $\gamma^* = \frac{|\varepsilon_{q^m}|}{|\varepsilon_{q^m}|+R-1} \in (1/R, 1)$ for $|\varepsilon_{q^m}| > 1$; $\gamma^* = 1$ is optimal under $\mathcal{I} = 1$. The optimal monetary policy regime is $\mathcal{I}^* = 0$, and the first-best allocation is not achievable with $\pi = 1$.*

The proof to this Proposition can be found in the appendix. The intuition behind the proof is as follows: We know from proposition 3.1 that the Friedman rule allows to achieve the first-best consumption level in the DM. Thus if aggregate labor supply is increasing in inflation or is independent from it ($\frac{\partial H}{\partial \gamma} \geq 0$, cases 1 and 3 from proposition 3.2) the Friedman rule must be the optimal monetary policy. Higher inflation would decrease consumption in the DM while increasing (or having no effect on) labor supply. If aggregate labor supply decreases in inflation ($\frac{\partial H}{\partial \gamma} < 0$), there is a policy tradeoff: Increasing the inflation rate reduces utility from DM consumption, but simultaneously reduces disutility from CM labor. This implies that the optimal inflation rate must lie above the Friedman rule. At the Friedman rule, the marginal costs of decreasing DM consumption are zero, but the benefits of decreasing the aggregate labor supply H are positive. Thus it is optimal to increase inflation above the Friedman rule. This is what happens in case 2 and in the model with $\mathcal{I} = 1$. By how much inflation can be increased above the Friedman rule to further increase welfare then depends on the elasticity of DM consumption and on the tax regime. With $\mathcal{I} = 0$, implementing deflation and thus higher DM consumption is relatively cheaper, so γ^* is increasing in the elasticity of DM consumption and approaching 1 as DM consumption becomes infinitely elastic. With $\mathcal{I} = 1$ however, running any deflationary policy in order to increase DM consumption is too costly, so $\gamma^* = 1$ in this case.

The welfare results also imply an ordering of the two monetary policy regimes. We know that for $\gamma = 1$, both regimes are equivalent in terms of allocation and we also know that if monetary policy is implemented over young buyers, $\gamma^* = 1$ is the optimal inflation rate. But this allocation is always feasible, but not optimal, if monetary policy is implemented over old buyers. Thus, we can conclude that the optimal monetary policy is to always set $\mathcal{I}^* = 0$, and to set $\gamma^* = 1/R$ for $|\varepsilon_{q^m}| \leq 1$ and to set $\gamma^* =\in (1/R, 1)$ for $|\varepsilon_{q^m}| > 1$.

Note that $\gamma = 1$ means that the return on money equals the return on intergenerational transfers which the social planner faces; note also that financing old-age CM consumption with fiat money implies intergenerational transfers, as only young agents are willing to sell goods against money. With $\pi = 1$, buyers are only able to compensate sellers with money for DM production, so unless the DM is completely shut down, sellers inevitably end up with money when they enter the second CM of their life. Setting $\gamma = 1$ leads to the correct prices for the two ways of generating CM consumption, i.e., intergenerational transfers and capital accumulation. On the other hand, setting $\gamma = \frac{1}{R}$ is the only way to reach the first-best consumption level in the DM. This analysis points out the fundamental policy tradeoff in our model: Efficiency in the DM requires a different money growth rate than efficiency in the CM, and the optimal policy choice depends on how much buyers value DM consumption, or more precisely, how elastic DM consumption is.

Haslag and Martin [2007] have shown that a constant money stock is typically optimal in an OLG model, independent of the Mundell-Tobin effect. We can confirm that $\gamma = 1$ is the optimal monetary policy for $\mathcal{I} = 1$, independent of all other parameters. However, we have also shown that a Mundell-Tobin effect still exists in this case, but it is running through the seller channel only. More generally, we can show that implementing monetary policy in a less costly way, namely by taxing agents only once they are old, allows to shut down the Mundell-Tobin effect completely for certain parameters, and that in these cases, the Friedman rule is the optimal monetary policy.[16]

The analysis so far shows that the first-best allocation is achieved automatically if capital is perfectly liquid, while the first-best consumption levels can be implemented at the Friedman rule in the case of perfectly illiquid capital, but at the cost of having agents work more in the first CM of their lives. Whether agents prefer this allocation over higher money growth rates in the case of perfectly illiquid capital depends on the way the Friedman rule is implemented and on the preferences of agents. But even if the Friedman rule is welfare maximizing in the case of perfectly illiquid capital, welfare is still strictly lower than in the case of perfectly liquid capital, i.e. below first best.

3.4 Equilibrium with partially liquid capital

We now turn to the case of partial relocation, which implies partially liquid capital. Uncertainty about relocation introduces a clear tradeoff between acquiring money or capital: Money provides insurance against the relocation shock, while capital offers a higher rate of return for $\phi_t/\phi_{t+1} > \frac{1}{R}$. At low inflation rates, acquiring money for insurance means only a small loss of return, thus making money relatively more attractive and depressing capital accumulation. At high inflation

[16]While this works nicely in our model, it would not do the trick in pure OLG models. The difference is that relocation occurs during the final stage of an agent's life in models such as Smith [2002] or Haslag and Martin [2007]. The reason that taxing the old is strictly cheaper in our model is that all agents know they have access to their capital when they have to pay the tax, and can thus fully pay the tax via capital investment. In pure OLG models, only non-relocated agents have access to their capital during the final stage of their life.

rates, acquiring money for insurance is really costly in terms of rate of return foregone, and thus capital accumulation becomes relatively more attractive. As we will see, this tradeoff adds a third channel through which capital accumulation is affected by the inflation rate. Since this channel only occurs when capital is partially liquid, we call it the *liquidity channel.*
For sellers, there is no uncertainty even with partially liquid capital. Thus, the results we found in equations (3.16) and (3.17) still hold. This implies two things: First, the seller's CM consumption is still unaffected by monetary policy and always at the first-best level; second, $p_t = \frac{R}{\phi_{t+1}}$ still holds.

Solving the buyers' lifetime problem (3.8) while making use of this result yields the following first-order conditions:

$$m_t : \quad \frac{\phi_t}{\phi_{t+1}} = \pi \frac{1}{R} u'\left(\frac{\phi_{t+1} m_t}{R}\right) + (1-\pi)\beta U'(\phi_{t+1} m_t + R(k_t^b - q_t^b) + (1-\mathcal{I})\tau^o) \tag{3.34}$$

$$k_t^b : \quad \frac{1}{\beta R} = \pi U'(R k_t^b) + (1-\pi) U'(\phi_{t+1} m_t + R(k_t^b - q_t^b) + (1-\mathcal{I})\tau^o) \tag{3.35}$$

$$q_t^b : \quad u'(q_t^b) = \beta R U'(\phi_{t+1} m_t + R(k_t^b - q_t^b) + (1-\mathcal{I})\tau^o). \tag{3.36}$$

We are now ready to define an equilibrium in the full model. Again we first derive the equilibrium when monetary policy is implemented over young buyers ($\mathcal{I} = 1$). Money market clearing $m = M$ and stationarity $\phi/\phi_{+1} = \gamma$ are identical to before but market clearing in the DM is now:

$$\pi q^m + (1-\pi) q^b = q^s. \tag{3.37}$$

With (3.37), (3.21), (3.24) and (3.25) and the definitions and first-order conditions derived above we can define a stationary equilibrium with partially liquid capital as a list of eleven variables $\{q^m$, q^b, q^s, x^b, x^m, x^s, $\phi_{+1}M$, k^b, h^b, k^s, $h^s\}$ solving:

$$\pi u'(q^m) + (1-\pi) u'(q^b) = \gamma R \tag{3.38}$$

$$\pi U'(x^m) + (1-\pi) U'(x^b) = \frac{1}{\beta R} \tag{3.39}$$

$$u'(q^b) = \beta R U'(x^b) \tag{3.40}$$

$$x^m = x^b + R(q^b - q^m) \tag{3.41}$$

$$x^s = x^* \tag{3.42}$$

$$k^{b,\mathcal{I}=1} = \frac{x^m}{R} \tag{3.43}$$

$$h^{b,\mathcal{I}=1} = \phi_{+1} M + k^b. \tag{3.44}$$

$$(3.45)$$

Aggregate labor supply and capital investments are given by:

$$K^{\mathcal{I}=1} = \frac{x^m}{R} + \frac{x^*}{R} - q^s \tag{3.46}$$

$$H^{\mathcal{I}=1} = \frac{x^m}{R} + q^m R + \frac{x^*}{R} - q^s. \tag{3.47}$$

With monetary policy implemented over old buyers ($\mathcal{I} = 0$) the only changes are in the labor supply and the capital investments of buyers and thus also in aggregate capital and labor supply.

$$h^{b,\mathcal{I}=0} = \gamma\phi_{+1}M + k^b \tag{3.48}$$

$$k^{b,\mathcal{I}=0} = \frac{x^m}{R} - q^m(\gamma - 1) \tag{3.49}$$

$$K^{\mathcal{I}=0} = \frac{x^m}{R} - q^m(\gamma - 1) + \frac{x^*}{R} - q^s \tag{3.50}$$

$$H^{\mathcal{I}=0} = \frac{x^m}{R} + \gamma q^m R - (\gamma - 1)q^m + \frac{x^*}{R} - q^s \tag{3.51}$$

Next, we are interpreting the equilibrium with a number of propositions. The proofs to all of them can be found in the appendix.

Proposition 3.4. *At the Friedman rule ($\gamma = \frac{1}{R}$), all DM trades are conducted with money, and the allocation is identical to an economy with $\pi = 1$. DM consumption is perfectly smoothed for relocated and non-relocated buyers and equal to the first-best level, i.e. $q^m = q^b = q^s = q^*$. CM consumption is also perfectly smoothed for relocated and non-relocated buyers and equal to first best $x^m = x^b = x^s = x^*$.*

The allocation under the Friedman rule achieves perfect insurance against the relocation shock and first best consumption in all markets. However, as we know from the last section, aggregate labor supply is above the first best in this equilibrium. All trades in the DM are made using money although capital would be accepted in some of them.

Proposition 3.5. *With inflation rates above the Friedman rule ($\gamma > \frac{1}{R}$), DM consumption is higher for non-relocated buyers and both consumption levels are below the first-best consumption level, i.e. $q^m < q^s < q^b < q^*$. CM consumption is higher for relocated buyers and above first-best consumption, while CM consumption for non-relocated buyers is below first best, i.e. $x^m > x^* > x^b$. Also, CM consumption for non-relocated buyers and DM consumption for all buyers decrease in inflation while CM consumption for relocated buyers increases in inflation, i.e. $\frac{\partial q^m}{\partial \gamma}, \frac{\partial q^b}{\partial \gamma}, \frac{\partial q^s}{\partial \gamma}, \frac{\partial x^b}{\partial \gamma} < 0$ and $\frac{\partial x^m}{\partial \gamma} > 0$. Total CM consumption $X = x^* + \pi x^m + (1 - \pi)x^b$ increases in inflation.*

The proposition shows that deviations from the Friedman rule introduce consumption risk for the agents and that their consumption deviates from first best in all markets. For relocated buyers, DM consumption is lower than for nonrelocated buyers, as they can only use money to purchase special goods. Therefore, their CM consumption has to be higher, as they still have the capital they accumulated during the DM. Nonrelocated buyers smooth their consumption more. They consume less than first-best in both markets because the low return on money (which they accumulate due to the ex-ante uncertainty about relocation) makes them unwilling to accumulate enough assets to purchase first-best consumption levels. In contrast to the model with full relocation, where buyer consumption in the CM was x^* and independent of inflation, inflation now affects CM consumption of relocated buyers (and total CM-consumption) positively.

The following two propositions describe the effects of inflation on total capital and the labor supply, in the two monetary policy regimes and at or above the Friedman rule, respectively.

Proposition 3.6. *With $\mathcal{I} = 1$, there is a Mundell-Tobin effect ($\frac{\partial K^{\mathcal{I}=1}}{\partial \gamma} > 0$) for all parameters. At the Friedman rule, an increase in inflation reduces aggregate labor supply ($\frac{\partial H^{\mathcal{I}=1}}{\partial \gamma} < 0$), but for sufficiently high inflation or liquidity of capital, a further increase in inflation increases aggregate labor supply.*

As in the model with fully illiquid capital, inflation always increases capital accumulation if monetary policy is implemented over young buyers. The effect is even stronger now, as the liquidity channel also makes buyers accumulate more capital instead of real balances if inflation increases in addition to the seller channel, which is still active. To see why the Mundell-Tobin effect and the effect on total labor supply don't always go in opposite directions anymore, we again look at market clearing in the CM in the model with taxes/transfers to young buyers (without loss of generality). Total CM-consumption is

$$X = x^* + \pi x^m + (1-\pi)x^b = K^{\mathcal{I}=1}R + \phi_{+1}M. \tag{3.52}$$

x^* is the sellers' CM consumption, and $\pi x^m + (1-\pi)x^b$ is the buyers' CM consumption. Total CM consumption is provided by capital investments $K^{\mathcal{I}=1}R$ and transfers $\phi_{+1}M$. Following the same steps as in the model with fully illiquid capital, total labor supply is $H^{\mathcal{I}=1} = X - K^{\mathcal{I}=1}(R-1)$. However, X is not independent of inflation anymore and thus the logic from before that an increase in inflation would always lead to a decrease in total labor is not valid anymore. From proposition 3.5 we know that X increases in inflation. Thus total labor supply still decreases over the Mundell-Tobin effect but increases over the effect of inflation on X. This is why there is no full correlation between the Mundell-Tobin effect and a negative effect of inflation on the labor supply anymore.

In a way, total labor supply now has an intensive (inflation shifts the financing mix of CM consumption from on the spot production to capital) and an extensive margin (inflation increases total consumption). One can show that the positive effect of inflation on CM consumption is zero at the Friedman rule. This is why at the Friedman rule, the connection between the Mundell-Tobin effect and a negative effect of inflation on total labor still holds.

Proposition 3.7. *With $\mathcal{I} = 0$, inflation affects labor supply and capital investment at the Friedman rule in the following way:*

1. *If $|\varepsilon_{q^m}|_{FR} < 1$: $\frac{\partial H^{\mathcal{I}=0}}{\partial \gamma} > 0$ and $\frac{\partial K^{\mathcal{I}=0}}{\partial \gamma} < 0$: Reverse Mundell-Tobin effect.*
2. *If $|\varepsilon_{q^m}|_{FR} > 1$: $\frac{\partial H^{\mathcal{I}=0}}{\partial \gamma} < 0$ and $\frac{\partial K^{\mathcal{I}=0}}{\partial \gamma} > 0$: Mundell-Tobin effect.*
3. *If $|\varepsilon_{q^m}|_{FR} = 1$: $\frac{\partial H^{\mathcal{I}=0}}{\partial \gamma} = 0$ and $\frac{\partial K^{\mathcal{I}=0}}{\partial \gamma} = 0$: No Mundell-Tobin effect.*

where $|\varepsilon_{q^m}|_{FR} = \frac{1}{\eta(q^*)}\xi$ and $\xi \in (1, \infty)$ with $\xi = 1$ if $\pi = 1$, $\xi \to \infty$ as $\pi \to 0$, and ξ monotonically decreasing in π at the Friedman rule. This implies that with a decrease in π, it is more likely that a Mundell-Tobin effect occurs at the Friedman rule, and if there is already a Mundell-Tobin effect, it becomes more pronounced.

Away from the Friedman rule, the conditions for a Mundell-Tobin effect and a negative effect on aggregate labor supply are not identical. $\frac{\partial K^{\mathcal{I}=0}}{\partial \gamma} > 0$ if $|\varepsilon_{q^m}| > \hat{\varepsilon}$ and $\frac{\partial H^{\mathcal{I}=0}}{\partial \gamma} < 0$ if $|\varepsilon_{q^m}| > \tilde{\varepsilon}$ where $\hat{\varepsilon} < 1 < \tilde{\varepsilon}$.

At the Friedman rule, the conditions for a Mundell-Tobin effect look identical to what we described in proposition 3.2 for $\pi = 1$ - there is a Mundell-Tobin effect for DM elasticity above 1, and a reverse Mundell-Tobin effect for DM elasticity below 1. However, note that the DM elasticity is now given by $\frac{1}{\eta(q^*)}\xi$ at the Friedman rule. Since $\xi = 1$ for $\pi = 1$ and ξ is decreasing in π, the proposition shows that with lower π, there are more values of $\eta(q^*)$ for which a Mundell-Tobin effect occurs at the Friedman rule; i.e., there are values for $\eta(q^*)$ for which a reverse Mundell-Tobin effect occurs for high values of π, but a Mundell-Tobin effect occurs for low values of π. If $\pi \to 0$, there is always a Mundell-Tobin effect at the Friedman rule. Further, if there is a Mundell-Tobin effect at the Friedman rule, the effect becomes stronger for lower π. The reason is the liquidity channel. At the Friedman rule, all DM trades are made using money and DM consumption is q^*. But if capital is relatively liquid (π is low) a lot of trades could be done using capital. The likelihood that buyers can use capital to pay in the DM is high. Thus if inflation is marginally above the Friedman rule and thus capital dominates money in terms of return, there is a strong incentive for buyers to substitute money for capital and massively increase capital holdings (and decrease labor as explained above). The lower π, the bigger this incentive, and the more likely that it dominates the agents' desire to smooth DM consumption which is captured by the risk-aversion coefficient. On the other hand if π is high, the incentive to increase capital is low at the Friedman rule if inflation marginally increases, because anyway most of the DM-trades are made using money, and the probability that buyers will be able to use capital to purchase goods in the DM is low.

Away from the Friedman rule, things are more complicated with partially liquid capital. In general, the Mundell-Tobin effect occurs when DM elasticity is above some threshold $\hat{\varepsilon} < 1$, but there is no simple representation of DM elasticity as a function of the coefficient of relative risk aversion in this case, so it is difficult to make general statements. Importantly, the threshold for whether inflation has a positive or a negative effect on aggregate labor supply is given by $\tilde{\varepsilon} > 1$, so away from the Friedman rule inflation does not generally have opposite effects on aggregate labor supply and aggregate capital accumulation. For $\hat{\varepsilon} < |\varepsilon_{q^m}| < \tilde{\varepsilon}$, a Mundell-Tobin effect coincides with a positive effect of inflation on aggregate labor supply.

Proposition 3.8. *For $\mathcal{I} = 1$, $\gamma^* = 1$ is the optimal money growth rate, independent of all other parameters. Furthermore, the Friedman rule is relatively more costly in this regime for low π.*

When monetary policy is implemented over young buyers, a constant money stock is optimal. This is not surprising, as we have shown in proposition 3.3 that a constant money stock is optimal in this monetary policy regime even if there is no uncertainty about relocation. With partially liquid capital, the Mundell-Tobin effect is generally stronger due to the liquidity channel, so there was no reason to expect a lower money growth rate to be optimal. Positive inflation rates are not optimal either, as for $\gamma > 1$, the additional distortions in consumption of relocated buyers are larger than benefits from increased capital accumulation. Further, the proposition also shows that running the Friedman rule is especially costly for low values of π. In fact, welfare at the Friedman rule is always the same independent of π, while welfare at $\gamma = 1$ is decreasing in π. The reason is that at the Friedman rule, all DM trades are made with money, which implies that a large share of the sellers' CM consumption is financed through intergenerational transfers, while at the first-best allocation, all CM consumption is financed by capital investment. Since for low levels of π, a large share of DM trades could be made using capital instead of money, the loss from running the Friedman rule is larger. For $\pi \to 0$, welfare at $\gamma = 1$ approaches the first-best.

Proposition 3.9. *For $\mathcal{I} = 0$ and $|\varepsilon_{q^m}|_{FR} \leq 1$, the optimal inflation rate is the Friedman rule,* *i.e.* $\gamma^* = 1/R$. *For* $|\varepsilon_{q^m}|_{FR} > 1$, *the optimal inflation rate is above the Friedman rule* $\gamma^* = \frac{|\varepsilon_{q^m}|}{|\varepsilon_{q^m}|+R-1} \in (1/R, 1)$. *In other words, for a given* π, *there is a threshold on risk aversion* ξ, *with the Friedman rule being the optimal monetary policy if and only if* $\eta(q^*) > \xi$, *and* ξ *is decreasing in* π. *When the optimal money growth rate is chosen, welfare is strictly higher with* $\mathcal{I} = 0$.

This proposition shows that higher liquidity of capital makes it less likely that the Friedman rule is the optimal monetary policy. In fact, when $\pi \to 0$, the Friedman rule is not optimal for any $\eta(q^*)$. This result directly follows from proposition 3.7. Due to the envelope theorem, an increase in inflation is welfare-increasing at the Friedman rule if it leads to a decrease in labor supply: such a decrease has a first-order effect on welfare, while the welfare losses from consumption in the DM and CM are only second order. Since a Mundell-Tobin effect coincides with a negative effect of inflation on labor supply at the Friedman rule, an increase in inflation is welfare-increasing if there is a Mundell-Tobin effect at the Friedman rule, which is the case for $|\varepsilon_{q^m}|_{FR} > 1$, with $|\varepsilon_{q^m}|_{FR} = \frac{1}{\eta(q^*)}\xi$. This shows that the liquidity channel makes it less likely that the Friedman rule is the optimal monetary policy. If the optimal monetary policy is not given by the Friedman rule, it is a function of the elasticity of DM consumption of relocated buyers, and it lies somewhere between the Friedman rule and a constant money stock. The result that monetary policy implementation over old buyers is better for welfare follows again from the fact that allocations coincide for $\gamma = 1$, so the optimal allocation for $\mathcal{I} = 1$ is always feasible with $\mathcal{I} = 0$, but generally not optimal.

As in the model with $\pi = 1$, the fundamental policy tradeoff is that setting $\gamma = \frac{1}{R}$ allows for efficiency in the DM, but misrepresents the cost of using money, and thus intergenerational transfers, to provide for CM consumption. While this is a relatively small issue if capital is illiquid and money is the only way to provide for DM consumption, the welfare loss from running the Friedman rule increases with the liquidity of capital, as most DM trades could be made with capital if it is relatively liquid.

Before we conclude, we want to note that proposition 3.9 shows that the premise behind Smith [2002] was correct: the Mundell-Tobin effect is enough to make deviations from the Friedman rule optimal. In our model, the Friedman rule is optimal if and only if there is no Mundell-Tobin effect at the Friedman rule, with higher liquidity of capital (i.e., a lower π) and lower risk aversion of buyers making it more likely that there is a Mundell-Tobin effect at the Friedman rule.

3.5 Conclusion

We have added a market which requires liquid assets to trade to an OLG model with relocation shocks, in order to study whether the Mundell-Tobin effect can make deviations from the Friedman rule optimal. We have shown that the Friedman rule is optimal if and only if there is no Mundell-Tobin effect at the Friedman rule, and that a Mundell-Tobin effect is more likely to occur if capital is relatively liquid, risk aversion of buyers is low, and if monetary policy is implemented by taxing the young. If the Friedman rule is not optimal, the optimal money growth rate lies somewhere between the Friedman rule and a constant money stock. While the Friedman rule allows for first-best consumption levels in the DM, it misrepresents the cost of using intergenerational transfers to provide for CM consumption during old age. These costs are correctly represented by a constant money stock. We have also shown that for any deflationary policy, taxing old agents is strictly better than taxing young agents when there is a productive investment opportunity in the economy.

Bibliography

Andrew Abel. Optimal monetary growth. *Journal of Monetary Economics*, 19:437–450, 1987.

Lukas Altermatt. Inside money, investment, and unconventional monetary policy. *University of Zurich, Department of Economics Working Paper No. 247*, 2019a.

Lukas Altermatt. Savings, asset scarcity, and monetary policy. *Journal of Economic Theory*, 182: 329–359, 2019b.

David Andolfatto, Aleksander Berentsen, and Christopher J. Waller. Monetary policy with asset-backed money. *Journal of Economic Theory*, 164:166–186, 2016.

S. Boragan Aruoba and Sanjay K. Chugh. Optimal fiscal and monetary policy when money is essential. *Journal of Economic Theory*, 145(5):1618–1647, 2010.

S. Boragan Aruoba and Randall Wright. Search, money, and capital: A neoclassical dichotomy. *Journal of Money, Credit and Banking*, 35(6):1085–1105, 2003. Part 2: Recent Developments in Monetary Economics.

S. Boragan Aruoba, Christopher Waller, and Randall Wright. Money and capital. *Journal of Monetary Economics*, 58:98–116, 2011.

Costas Azariadis and Bruce D. Smith. Private information, money, and growth: Indeterminacy, fluctuations, and the mundell-tobin effect. *Journal of Economic Growth*, 1:309–332, 1996.

Joydeep Bhattacharya, Joseph Haslag, and Steven Russell. The role of money in two alternative models: When is the friedman rule optimal, and why? *Journal of Monetary Economics*, 52: 1401–1433, 2005.

Scott Freeman. Resolving differences over the optimal quantity of money. *Journal of Money, Credit and Banking*, 25(4):801–811, 1993.

Milton Friedman. The optimum quantity of money. *The Optimum Quantity of Money and Other Essays*, pages 1–50, 1969.

Athanasios Geromichalos and Lucas Herrenbrueck. The liquidity-augmented model of macroeconomic aggregates. *Simon Fraser University, Department of Economics Working Paper 17-16*, 2017.

Pedro Gomis-Porqueras, Stella Huangfu, and Hongfei Sun. The role of search frictions in the long-run relationships between inflation, unemployment and capital. *European Economic Review*, 123, 2020.

Jean Michel Grandmont and Yves Younes. On the efficiency of a monetary equilibrium. *Review of Economic Studies*, 40:149–165, 1973.

Joseph H. Haslag and Antoine Martin. Optimality of the friedman rule in an overlapping generations model with spatial separation. *Journal of Money, Credit and Banking*, 39(7):1741–1758, 2007.

Ryoji Hiraguchi. Optimal monetary policy in an overlapping generations model with search theoretic monetary exchange. *The B.E. Journal of Theoretical Economics*, 17(2), 2017.

Samuel Huber and Jaehong Kim. An overlapping generations model for monetary policy analysis. *European Economic Review*, 125, 2020.

Ricardo Lagos and Guillaume Rocheteau. Money and capital as competing media of exchange. *Journal of Economic Theory*, 142:247 – 258, 2008.

Ricardo Lagos and Randall Wright. A unified framework for monetary theory and policy analysis. *Journal of Political Economy*, 113 (3):463–484, 2005.

Robert E. Lucas and Nancy L. Stokey. Money and interest in a cash-in-advance economy. *Econometrica*, 55(3):491–513, 1987.

Robert Mundell. Inflation and real interest. *Journal of Political Economy*, 71(3):280–283, 1963.

Daniel Sanches and Stephen Williamson. Money and credit with limited commitment and theft. *Journal of Economic Theory*, 145:1525–1549, 2010.

Stephanie Schmitt-Grohé and Martín Uribe. The Optimal Rate of Inflation. In Benjamin M. Friedman and Michael Woodford, editors, *Handbook of Monetary Economics*, volume 3 of *Handbook of Monetary Economics*, chapter 13, pages 653–722. Elsevier, 2010.

Stacey Schreft and Bruce D. Smith. The conduct of monetary policy with a shrinking stock of government debt. *Journal of Money, Credit and Banking*, 34:848–882, 2002.

Stacey L. Schreft and Bruce D. Smith. Money, banking, and capital formation. *Journal of Economic Theory*, 73:157–182, 1997.

Bruce D. Smith. Monetary policy, banking crises, and the friedman rule. *American Economic Review Papers and Proceedings*, 92(2):128–134, 2002.

Bruce D. Smith. Taking intermediation seriously. *Journal of Money, Credit and Banking*, 35: 1319–1357, 2003.

James Tobin. Money and economic growth. *Econometrica*, 33(4):671–684, 1965.

Robert M. Townsend. Models of money with spatially separated agents. In John Kareken and Neil Wallace, editors, *Models of Monetary Economies*, pages 265–303. Minneapolis: Federal Reserve Bank of Minneapolis, 1980.

Robert M. Townsend. Economic organization with limited communication. *The American Economic Review*, 77(5):954–971, 1987.

Venky Venkateswaran and Randall Wright. Pledgability and liquidity: A new monetarist model of financial and macroeconomic activity. *NBER Macroeconomics Annual*, 28:227–270, 2013.

Neil Wallace. The overlapping generations model of fiat money. *Models of Monetary Economies, Federal Reserve Bank of Minneapolis*, pages 49–82, 1980.

Laurence Weiss. The effects of money supply on economic welfare in the steady state. *Econometrica*, 48(3):565–576, 1980.

Stephen Williamson. Liquidity, monetary policy, and the financial crisis: A new monetarist approach. *American Economic Review*, 102 (6):2570–2605, 2012.

Randall Wright, Sylvia Xiaolin Xiao, and Yu Zhu. Frictional capital reallocation i: Ex ante heterogeneity. *Journal of Economic Dynamics & Control*, 89:100–116, 2018.

Randall Wright, Sylvia Xiao, and Yu Zhu. Frictional capital reallocation ii: Ex post heterogeneity. *mimeo*, 2019.

Tao Zhu. An overlapping-generations model with search. *Journal of Economic Theory*, 142(1): 318–331, 2008.

Appendix C

C.1 Proof of Proposition 3.2

Proof. We begin with the Mundell-Tobin effect in both models. When monetary policy is implemented over young buyers ($\mathcal{I} = 0$) total capital investment is given by:

$$K^{\mathcal{I}=1} = \frac{2x^*}{R} - q^m \tag{3.27}$$

DM-consumption shows up because it indirectly affects the sellers' capital accumlation. Since q^m is decreasing in inflation, we always get a Mundell-Tobin effect when monetary policy is implemented over young buyers:

$$\frac{\partial K^{\mathcal{I}=1}}{\partial \gamma} = -\frac{\partial q^m}{\partial \gamma} > 0 \tag{C.1}$$

When monetary policy is implemented over old buyers ($\mathcal{I} = 0$) total capital investment is:

$$K^{\mathcal{I}=0} = \frac{2x^*}{R} - q^m(\gamma - 1) - q^m = \frac{2x^*}{R} - q^m\gamma \tag{3.31}$$

with the first derivative:

$$\frac{\partial K^{\mathcal{I}=0}}{\partial \gamma} = -(\gamma\frac{\partial q^m}{\partial \gamma} + q^m) = \frac{u'(q^m)}{-u''(q^m)} - q^m = q^m\big(\frac{1}{\eta(q^m)} - 1\big). \tag{C.2}$$

Furthermore, $\frac{1}{\eta(q^m)} = |\varepsilon_{q^m}|$, since

$$|\varepsilon_{q^m}| = -\frac{dq^m/q^m}{d\gamma/\gamma} = -\frac{\gamma}{q^m}\frac{\partial q^m}{\partial \gamma} = -\frac{u'(q^m)}{q^m u''(q^m)}. \tag{C.3}$$

Thus $\frac{\partial K^{\mathcal{I}=0}}{\partial \gamma} > 0$ (we have a Mundell-Tobin effect) if $|\varepsilon_{q^m}| = \frac{1}{\eta(q^m)} > 1$.

Next we turn to the effects on total labor supply. In both monetary policy regimes, total labor supply is the sum of capital investments K and the work of buyers to acquire real balances. If monetary policy is instead implemented over the young buyers ($\mathcal{I} = 1$), total labor supply is:

$$H^{\mathcal{I}=1} = \gamma q^m R + \frac{x^*}{R} - q^m R(\gamma - 1) + \frac{x^*}{R} - q^m = \frac{2x^*}{R} + q^m(R - 1) \tag{3.28}$$

Buyers acquire real balances $\phi M = \gamma q^m R$ and get a transfer (if $\gamma > 1$) of $\tau = q^m R(\gamma - 1)$). So in this case the wealth effects of inflation on the holdings of real balances are canceled out and the effect of inflation on total labor supply must be negative:

$$\frac{\partial H^{\mathcal{I}=1}}{\partial \gamma} = \frac{\partial q^m}{\partial \gamma}(R-1) < 0 \tag{C.4}$$

With monetary policy implemented over old buyers ($\mathcal{I} = 0$) total labor supply in the CM is in equilibrium:

$$H^{\mathcal{I}=0} = \gamma q^m R + \frac{x^*}{R} - q^m(\gamma - 1) + \frac{x^*}{R} - q^m = K^{\mathcal{I}=0} + \gamma q^m R = \frac{2x^*}{R} + (R-1)\gamma q^m \tag{3.32}$$

Total labor supply is the sum of buyer real balances $\gamma q^m R$ and total capital investments. The effects of real balances on $K^{\mathcal{I}=0}$ and the real balance holdings $\gamma q^m R$ simplify to $(R-1)\gamma q^m$. Thus the sign of the derivative of γq^m, which is determined by $\eta(q^m)$, will also determine the sign of the derivative of total labor supply:

$$\frac{\partial H^{\mathcal{I}=0}}{\partial \gamma} = (R-1)\left(\frac{\partial q}{\partial \gamma}\gamma + q^m\right) = q^m(R-1)\big(1 - \frac{1}{\eta(q^m)}\big) \tag{C.5}$$

Thus we must have $\frac{\partial K^{\mathcal{I}=0}}{\partial \gamma} > 0$, $\frac{\partial H^{\mathcal{I}=0}}{\partial \gamma} < 0$ for $|\varepsilon_{q^m}| = \frac{1}{\eta(q^m)} > 1$, and $\frac{\partial K^{\mathcal{I}=0}}{\partial \gamma} < 0$, $\frac{\partial H^{\mathcal{I}=0}}{\partial \gamma} > 0$ for $|\varepsilon_{q^m}| = \frac{1}{\eta(q^m)} < 1$.

■

C.2 Proof of Proposition 3.3

Proof. Welfare of a representative generation with fully illiquid capital can be written as

$$V^g = -H + u(q^m) - q^m + 2\beta U(x^*) \tag{C.6}$$

Because CM consumption is independent of inflation for $\pi = 1$, inflation affects welfare through DM consumption and aggregate labor supply:

$$\frac{\partial V^g}{\partial \gamma} = -\frac{\partial H}{\partial \gamma} + \frac{\partial q^m}{\partial \gamma}(u'(q^m) - 1) \tag{C.7}$$

In cases 1 and 3 from proposition 3.2, the aggregate labor supply rises in inflation or is independent from it ($\frac{\partial H}{\partial \gamma} \geq 0$). Thus, the Friedman rule must be the optimal monetary policy since this maximizes welfare in the DM ($q^m = q^*$) from (3.19). Higher inflation would decrease consumption in the DM while weakly increasing labor supply.[1]

In case 2 and with $\mathcal{I} = 1$, aggregate labor supply decreases in inflation ($\frac{\partial H}{\partial \gamma} < 0$). Thus, the optimal inflation rate must lie above the Friedman rule due to the envelope theorem.

In case 2 under $\mathcal{I} = 0$, $\frac{\partial H}{\partial \gamma}$ given by (C.5) and optimal inflation rate $\gamma^* < 1$ solves

$$-q^m(R-1) - \frac{\partial q^m}{\partial \gamma}\gamma^*(R-1) + \frac{\partial q^m}{\partial \gamma}(\gamma^* R - 1) = 0$$

$$\leftrightarrow$$

$$-\frac{\partial q^m}{\partial \gamma}\frac{\gamma^*}{q^m} = |\varepsilon_{q^m}| = \frac{(R-1)\gamma^*}{1-\gamma^*} \tag{C.8}$$

The interior solution solves

$$\gamma^* = \frac{|\varepsilon_{q^m}|}{|\varepsilon_{q^m}| + R - 1} \quad \in (1/R, 1) \tag{C.9}$$

If monetary policy is implemented over young buyers ($\mathcal{I} = 1$), $\frac{\partial H}{\partial \gamma}$ is given by (C.4) and optimal inflation rate γ^* solves

$$-\frac{\partial q}{\partial \gamma}(R-1) + \frac{\partial q}{\partial \gamma}(\gamma^* R - 1) = 0 \tag{C.10}$$

Thus $\gamma^* = 1$.

■

C.3 Proof of Proposition 3.4

Proof. We proof this by contradiction using the first four optimality conditions (3.38) to (3.41) which we repeat for convenience.

$$\pi u'(q^m) + (1-\pi)u'(q^b) = \gamma R \tag{3.38}$$

[1]For inflation rates below the Friedman rule, our derivation of results is incorrect, because we assumed $\gamma \geq \frac{1}{R}$. It looks like further decreasing inflation is welfare-increasing if $|\varepsilon_{q^m}| < 1$, but this is incorrect, as inflation below the Friedman rule leads to a regime switch where nobody accumulates capital. This clearly reduces aggregate welfare. Thus, $\gamma = \frac{1}{R}$ is a corner solution for $|\varepsilon_{q^m}| < 1$.

$$\pi U'(x^m) + (1-\pi)U'(x^b) = \frac{1}{\beta R} \tag{3.39}$$

$$u'(q^b) = \beta R U'(x^b) \tag{3.40}$$

$$x^m = x^b + R(q^b - q^m) \tag{3.41}$$

At the Friedman rule (3.38) becomes:

$$\pi u'(q^m) + (1-\pi)u'(q^b) = 1$$

Suppose $q^b > q^m$. From (3.41) this implies $x^m > x^b$. From the decreasing marginal utility of the utility functions and the weighted average formulation of the first and the second condition we must have $u'(q^m) > 1$ and $u'(q^b) < 1$ and $U'(x^m) > \frac{1}{\beta R}$ and $U'(x^b) < \frac{1}{\beta R}$. But this violates (3.40), the left-hand side would be < 1 and the right-hand side > 1. The opposite contradiction follows for assuming $q^b < q^m$. Thus we must have $q^b = q^m$ and $x^m = x^b$ and then $q^b = q^m = q^*$ and $x^m = x^b = x^*$ follow from (3.38) and (3.39). From $q^b = q^s$, it follows that all DM trades are made with money, and from that it follows that the allocation must be identical to an economy with $\pi = 1$. ■

C.4 Proof of Proposition 3.5

Proof. We first show that for $\gamma > 1/R$ we must have $q^* > q^b > q^m$ and $x^m > x^* > x^b$. The steps are the same as in the previous proof. Suppose that $q^b = q^m$ and $x^m = x^b$ from (3.41). This implies $u'(q) > 1$ from (3.38) and $\beta R U'(x) = 1$. Thus it violates (3.40). Similarly $q^b < q^m$ and $x^m < x^b$ imply $u'(q^b) > \gamma R > 1$ and $U'(x^b) < \frac{1}{\beta R}$ which also violates (3.40). Thus only $q^b > q^m$ and $x^m > x^b$ is possible, implying $u'(q^b) < \gamma R$ and $U'(x^b) > \frac{1}{\beta R}$. In this case (3.40) can hold if $u'(q^b) \in (1, \gamma R)$ which also implies $q^b < q^*$. Since $q^m < q^b$, q^m must also be below q^* and $x^m > x^*$ and $x^b < x^*$ follows from $U'(x^b) > \frac{1}{\beta R}$ and $U'(x^m) < \frac{1}{\beta R}$.

To find the effects of inflation on the consumption levels we differentiate (3.38) to (3.41) with respect to γ.

$$\begin{aligned}
\pi u''(q^m)\frac{\partial q^m}{\partial \gamma} + (1-\pi)u''(q^b)\frac{\partial q^b}{\partial \gamma} &= R \\
\pi U''(x^m)\frac{\partial x^m}{\partial \gamma} + (1-\pi)U''(x^b)\frac{\partial x^b}{\partial \gamma} &= 0 \\
u''(q^b)\frac{\partial q^b}{\partial \gamma} &= \beta R U''(x^b)\frac{\partial x^b}{\partial \gamma} \\
\frac{\partial x^m}{\partial \gamma} &= \frac{\partial x^b}{\partial \gamma} + R(\frac{\partial q^b}{\partial \gamma} - \frac{\partial q^m}{\partial \gamma})
\end{aligned} \tag{C.11}$$

Solving this for the partial effects yields

$$\frac{\partial q^b}{\partial \gamma} = A\frac{\partial x^b}{\partial \gamma} < 0 \tag{C.12}$$

$$\frac{\partial x^m}{\partial \gamma} = -B\frac{\partial x^b}{\partial \gamma} > 0 \tag{C.13}$$

$$\frac{\partial q^m}{\partial \gamma} = \frac{(RA+1+B)}{R}\frac{\partial x^b}{\partial \gamma} < 0 \tag{C.14}$$

$$\frac{\partial x^b}{\partial \gamma} = \frac{R}{C} < 0, \tag{C.15}$$

where $A = \frac{\beta RU''(x^b)}{u''(q^b)}$ and $B = \frac{(1-\pi)U''(x^b)}{\pi U''(x^m)}$ are positive and $C = \pi u''(q^m)\frac{RA+1+B}{R} + (1-\pi)u''(q^b)A$ is negative because $u(\cdot)$ and $U(\cdot)$ are strictly concave. The effect on seller DM consumption q^s must also be negative because it is a weighted average of consumption of relocated and non-relocated buyers from (3.37).

The partial effect of inflation on aggregate CM-consumption $X = x^* + \pi x^m + (1-\pi)x^b$ is:

$$\frac{\partial X}{\partial \gamma} = \pi\frac{\partial x^m}{\partial \gamma} + (1-\pi)\frac{\partial x^b}{\partial \gamma} = -(\pi(1+B)-1)\frac{\partial x^b}{\partial \gamma} \tag{C.16}$$

which is positive since $\frac{\partial x^b}{\partial \gamma} < 0$ and $\pi(1+B) > 1$ for inflation rates above the Friedman rule as we show in the proof of proposition 3.6. ■

C.5 Proof of Proposition 3.6

Proof. With $\mathcal{I} = 1$, aggregate capital investment and labor supply are in equilibrium:

$$K^{\mathcal{I}=1} = \frac{x^m}{R} + \frac{x^*}{R} - q^s \tag{3.46}$$

$$H^{\mathcal{I}=1} = \frac{x^m}{R} + q^m R + \frac{x^*}{R} - q^s = K^{\mathcal{I}=1} + q^m R \tag{3.47}$$

Deriving $K^{\mathcal{I}=1}$ with respect to inflation γ yields

$$\frac{\partial K^{\mathcal{I}=1}}{\partial \gamma} = -\frac{\partial q^m}{\partial \gamma}\mathcal{X}_1 > 0, \tag{C.17}$$

with $\mathcal{X}_1 = \frac{B+RA+\pi(1+B)}{B+RA+1}$.

Already a visual inspection of (3.46) tells us that there must be a Mundell-Tobin effect. From the proof of proposition 3.5, x^m increases with inflation and q^m and q^b and thus also $q^s = \pi q^m + (1-\pi)q^b$ decrease with inflation. The derivative confirms this, as all terms in $\mathcal{X}_1$ are positive.

From (3.47), the effect of inflation on aggregate labor supply in the first CM is the sum of a positive Mundell-Tobin effect and a negative effect over the real money holdings of buyers $q^m R$. This can be written as:

$$\frac{\partial H^{\mathcal{I}=1}}{\partial \gamma} = -\frac{\partial q^m}{\partial \gamma}\mathcal{X}_1 + R\frac{\partial q^m}{\partial \gamma} = \frac{\partial q^m}{\partial \gamma}(R - \mathcal{X}_1), \tag{C.18}$$

so the effect of an increase in inflation on aggregate labor supply depends on $R \lesseqgtr \mathcal{X}_1$.

From the definition of B in the proof to proposition 3.5, we know

$$\pi(1+B) = \pi + (1-\pi)\frac{U''(x^b)}{U''(x^m)} \geq 1,$$

since $x^m \geq x^b$ and $U'''(x) > 0$.[2] This implies $\mathcal{X}_1 \geq 1$. $\mathcal{X}_1 = 1$ if either $\pi = 1$ (capital is fully illiquid) or if $x^m = x^b = x^*$ (at the Friedman rule), and strictly larger otherwise. From the partial derivatives, increasing inflation from the Friedman rule increases the spread between DM-consumption x^m/x^b. Therefore $\pi(1+B)$ and $\mathcal{X}_1$ must rise with γ, while they are decreasing in π. Thus, the effect of inflation on aggregate labor is always negative at the Friedman rule. Away from the Friedman rule, higher inflation and higher liquidity of capital make it more likely that there is a positive effect of inflation on aggregate labor supply.

■

C.6 Proof of Proposition 3.7

Proof. With $\mathcal{I} = 0$, aggregate capital investment and labor supply are in equilibrium:

$$K^{\mathcal{I}=0} = \frac{x^m}{R} + \frac{x^*}{R} - q^s - q^m(\gamma - 1) \tag{3.50}$$

$$H^{\mathcal{I}=0} = \frac{x^m}{R} + \gamma q^m R - (\gamma-1)q^m + \frac{x^*}{R} - q^s = K^{\mathcal{I}=0} + \gamma q^m R \tag{3.51}$$

Differentiating aggregate capital investment with respect to inflation yields:

$$\frac{\partial K^{\mathcal{I}=0}}{\partial \gamma} = -\Big(\frac{\partial q^m}{\partial \gamma}\gamma\mathcal{X}_0 + q^m\Big), \tag{C.19}$$

with $\mathcal{X}_0 = \frac{\pi(1+B)-1}{\gamma(RA+1+B)} + 1$. $\mathcal{X}_0 = 1$ for $\pi = 1$ (since $B = 0$ for $\pi = 1$) and at the Friedman rule (since $B = \frac{1-\pi}{\pi}$ at the FR). The condition for a Mundell-Tobin effect is

$$-\frac{\partial q^m}{\partial \gamma} > \frac{q^m}{\gamma \mathcal{X}_0} \tag{C.20}$$

[2] $U''(x^b) < U''(x^m)$, but since both second derivatives are negative ($U''(x^b), U''(x^m) < 0$), $U''(x^b)/U''(x^m) > 1$.

$$|\varepsilon_{q^m}| > \frac{1}{\mathcal{X}_0} = \hat{\varepsilon} \tag{C.21}$$

using the definition of the elasticity of DM-consumption with respect to inflation: $|\varepsilon_{q^m}| = -\frac{\gamma}{q^m}\frac{\partial q^m}{\partial \gamma}$.

At the Friedman rule $\mathcal{X}_0 = 1$, so there is a Mundell-Tobin effect if $|\varepsilon_{q^m}|_{FR} > 1$. Using (C.14) and (C.15) with $q^m = q^b = q^*$ and $x^m = x^b = x^*$, the derivative of q^m with respect to inflation and the elasticity are:

$$-\frac{\partial q^m}{\partial \gamma}|_{FR} = \frac{1}{u''(q^*)}\frac{R(\pi RA + 1)}{\pi(1 + RA)} \tag{C.22}$$

$$|\varepsilon_{q^m}|_{FR} = -\frac{u'(q^*)}{q^* u''(q^*)}\xi, \tag{C.23}$$

with $\xi = \frac{1+\pi RA}{\pi(1+RA)}$. Thus, there is a Mundell-Tobin effect at the Friedman rule if:

$$-\frac{u'(q^*)}{q^* u''(q^*)}\xi > 1$$

$$\Rightarrow -\frac{u'(q^*)}{q^* u''(q^*)} > \frac{1}{\xi}. \tag{C.24}$$

$\xi \to 1$ for $\pi \to 1$, and $\xi \to \infty$ for $\pi \to 0$, and it can easily be shown that ξ is monotonically decreasing in π at the Friedman rule. This implies that for a given risk aversion, a Mundell-Tobin effect is more likely to occur at the Friedman rule for lower π.

Combining (3.51) and (C.2) the derivative of the aggregate labor supply with respect to inflation is:

$$\frac{\partial H^{\mathcal{I}=0}}{\partial \gamma} = (R-1)(\frac{\partial q^m}{\partial \gamma}\gamma + q^m) - \frac{\partial q^m}{\partial \gamma}(\mathcal{X}_0 - 1). \tag{C.25}$$

This is negative if

$$-\frac{\partial q^m}{\partial \gamma} > \frac{q^m(R-1)}{\gamma(R-1) - (\mathcal{X}_0 - 1)} \tag{C.26}$$

$$|\varepsilon_{q^m}| > \frac{\gamma(R-1)}{\gamma(R-1) - (\mathcal{X}_0 - 1)} = \tilde{\varepsilon}, \tag{C.27}$$

so there is a negative effect of inflation on aggregate labor supply for $\gamma(R-1) > \mathcal{X}_0 - 1$. Since $\mathcal{X}_0 = 1$ at the Friedman rule, $\hat{\varepsilon}|_{FR} = \tilde{\varepsilon}|_{FR} = 1$, so the conditions coincide, and a Mundell-Tobin effect (reverse Mundell-Tobin effect) always implies a negative (positive) effect of inflation on aggregate labor supply. Away from the Friedman rule, $\mathcal{X}_0 > 1$, so $\hat{\varepsilon} < 1$ while $\tilde{\varepsilon} > 1$, and thus there is a range of values for $|\varepsilon_{q^m}|$ for which a Mundell-Tobin effect comes along with a positive effect of inflation on aggregate labor supply.

■

C.7 Proof of Proposition 3.8

Proof. With monetary policy implemented over young buyers expected welfare of a representative generation is given by:

$$
\begin{aligned}
V^{\mathcal{I}=1} &= -H^{\mathcal{I}=1} + \pi(u(q^m) + \beta U(x^m)) + (1-\pi)(u(q^b) + \beta U(x^b)) - q^s + \beta U(x^*) && \text{(C.28)} \\
&= -\frac{x^m}{R} - q^m R - \frac{x^*}{R} + \pi(u(q^m) + \beta U(x^m)) + (1-\pi)(u(q^b) + \beta U(x^b)) + \beta U(x^*)
\end{aligned}
$$

Differentiating (C.28) with respect to inflation and replacing $u'(q^m), U'(x^m)$ and $u'(q^b)$ using the optimality conditions (3.38), (3.39) and (3.41) yields:

$$
\frac{\partial V^{\mathcal{I}=1}}{\partial \gamma} = -R\frac{\partial q^m}{\partial \gamma} + \gamma R \frac{\partial q^m}{\partial \gamma} - (1-\pi)\beta R U'(x^b)\Big(\frac{\partial q^m}{\partial \gamma} - \frac{\partial q^b}{\partial \gamma} + \frac{\frac{\partial x^m}{\partial \gamma} - \frac{\partial x^b}{\partial \gamma}}{R}\Big)
$$

The last bracket is equal to zero from (C.11), and we obtain the identical expression as in the model with fully illiquid capital, (C.10):

$$
\frac{\partial V^{\mathcal{I}=1}}{\partial \gamma} = \frac{\partial q^m}{\partial \gamma} R(\gamma - 1) \tag{C.29}
$$

At the Friedman rule this expression must be positive since $\frac{\partial q^m}{\partial \gamma} < 0$ from (C.14) (and from proposition 3.7 the marginal effect on the labor supply is negative). From there expected welfare increases in inflation for $\gamma < 1$ and decreases for $\gamma > 1$. Thus the unique optimum must be $\gamma^* = 1$ independent of π and the other parameters.

To show that the Friedman rule is relatively more costly at lower levels of π, we show that $V^{\mathcal{I}=1}(\gamma^*) - V^{\mathcal{I}=1}(\gamma^{FR})$, the difference in expected welfare under optimal monetary policy ($\gamma^* = 1$) and the Friedman rule ($\gamma = 1/R$), decreases in π. From proposition 3.4, welfare at the Friedman rule is given by:

$$
V^{\mathcal{I}=1}(\gamma^{FR}) = -\frac{2x^*}{R} - q^* R + u(q^*) + 2\beta U(x^*) \tag{C.30}
$$

which is independent of π. Thus to show that $V^{\mathcal{I}=1}(\gamma^*) - V^{\mathcal{I}=1}(\gamma^{FR})$ decreases in π it is sufficient to show that expected welfare under optimal monetary policy is decreasing in π. From (C.28) $V^{\mathcal{I}=1}(\gamma^*)$ is given by:

$$
V^{\mathcal{I}=1}(\gamma^*) = -\frac{x^m}{R} - q^m R - \frac{x^*}{R} + \pi(u(q^m) + \beta U(x^m)) + (1-\pi)(u(q^b) + \beta U(x^b)) + \beta U(x^*) \tag{C.31}
$$

where all optimality conditions (3.38) to (3.41) hold and (3.38) is evaluated at $\gamma^* = 1$. Since (C.31) is evaluated at the optimum we can ignore the indirect effects of π on the variables and directly take the partial derivative of (C.31) with respect to π.

$$
\frac{\partial V^{\mathcal{I}=1}(\gamma^*)}{\partial \pi} = u(q^m) + \beta U(x^m) - (u(q^b) + \beta U(x^b)) < 0. \tag{C.32}
$$

Changes in expected welfare at γ^* through π reflect differences in the utility of consumption as a relocated and a non-relocated buyer. If utility of consumption as a relocated buyer is higher, expected welfare would rise with π and vice versa. Since $q^b > q^m$ and $x^m > x^b$ it is not a priori clear where utility is higher. However, since for a non-relocated buyer the relocated allocation $\{q^m, x^m\}$ is also feasible but not chosen, it must be that utility of non-relocated buyers is higher or $u(q^b) + \beta U(x^b) > u(q^m) + \beta U(x^m)$. Thus $\frac{\partial V^{\mathcal{I}=1}(\gamma^*)}{\partial \pi} < 0$, implying that expected welfare at the optimal monetary policy is decreasing in π. In turn, this shows that the welfare loss of running the Friedman rule is decreasing in π.

■

C.8 Proof of Proposition 3.9

Proof. Expected welfare when monetary policy is implemented over old buyers is given by:

$$\begin{aligned} V^{\mathcal{I}=0} &= -H^{\mathcal{I}=0} + \pi(u(q^m) + \beta U(x^m)) + (1-\pi)(u(q^b) + \beta U(x^b)) - q^s + \beta U(x^*) \qquad \text{(C.33)} \\ &= -\frac{x^m}{R} - q^m R\gamma + (\gamma - 1)q^m - \frac{x^*}{R} + \pi(u(q^m) + \beta U(x^m)) + (1-\pi)(u(q^b) + \beta U(x^b)) + \beta U(x^*) \end{aligned}$$

Differentiating (C.33) with respect to inflation and replacing $u'(q^m), U'(x^m)$ and $u'(q^b)$ using the optimality conditions (3.38), (3.39) and (3.41) yields:

$$\frac{\partial V^{\mathcal{I}=0}}{\partial \gamma} = \frac{\partial q^m}{\partial \gamma}(\gamma - 1) - (R-1)q^m - (1-\pi)\beta R U'(x^b)\Big(\frac{\partial q^m}{\partial \gamma} - \frac{\partial q^b}{\partial \gamma} + \frac{\frac{\partial x^m}{\partial \gamma} - \frac{\partial x^b}{\partial \gamma}}{R}\Big)$$

The last bracket is zero from (C.11) and we obtain the identical expression as in the model with fully illiquid capital, (C.8):

$$\frac{\partial V^{\mathcal{I}=0}}{\partial \gamma} = \frac{\partial q^m}{\partial \gamma}(\gamma - 1) - (R-1)q^m \qquad \text{(C.34)}$$

Since $\frac{\partial q^m}{\partial \gamma} < 0$ an increase in inflation can only be welfare improving if there is deflation ($\gamma < 1$). When is a deviation from the Friedman rule welfare improving? Evaluating (C.34) at the Friedman rule:

$$\frac{\partial V^{\mathcal{I}=0}}{\partial \gamma}|_{FR} = q^*(R-1)\Big(|\varepsilon_{q^m}|_{FR} - 1\Big) \qquad \text{(C.35)}$$

Thus an inflation rate above the Friedman rule, $\gamma^* > 1/R$ is optimal if $|\varepsilon_{q^m}|_{FR} - 1$ holds, which is the same condition as for a Mundell-Tobin effect and a negative effect on aggregate labor supply.

If $|\varepsilon_{q^m}|_{FR} - 1$ holds, the optimal inflation rate $\gamma^* > 1/R$ is given by (C.34) set to 0 which is exactly the same expression as for the case of fully illiquid capital (C.8)

$$(1 - \gamma^*) - \frac{\partial q^m}{\partial \gamma} = (R-1)q^m \qquad \text{(C.36)}$$

$$\leftrightarrow \tag{C.37}$$

$$-\frac{\partial q^m}{\partial \gamma}\frac{\gamma^*}{q^m} = |\varepsilon_{q^m}| = \frac{(R-1)\gamma^*}{1-\gamma^*}. \tag{C.38}$$

It can easily be seen that the right-hand-side of this expression equals 1 at the Friedman rule, ∞ for $\gamma = 1$, and is strictly increasing in γ $\forall \gamma \in \{\frac{1}{R}, 1\}$. From the proof to proposition 3.7, we know that $|\varepsilon_{q^m}|_{FR} = \frac{1}{\eta(q^*)}\xi$. Thus, the Friedman rule is optimal if and only if $\eta(q^*) > \xi$. Since ξ is monotonically decreasing in π and $\xi \to \infty$ for $\pi \to 0$, higher liquidity of capital (lower π) makes it less likely for the FR to be optimal, and the FR is never optimal when capital is perfectly liquid.

With partially liquid capital, the optimal inflation rate γ^* is still given by (C.9):

$$\gamma^* = \frac{|\varepsilon_{q^m}|}{|\varepsilon_{q^m}| + R - 1,} \quad \in (1/R, 1) \tag{C.9}$$

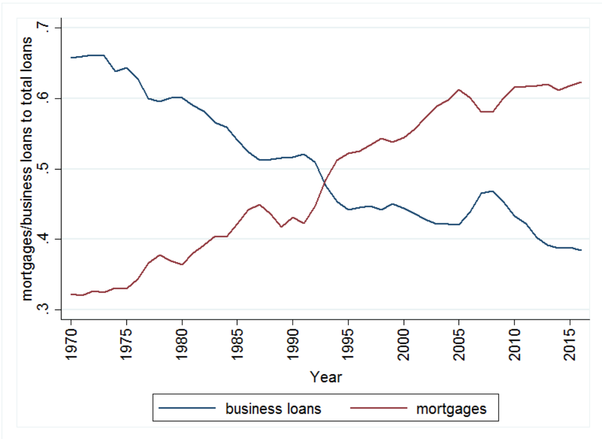
but with $|\varepsilon_{q^m}|$ now also being a function of π.

The result that welfare is higher with $\mathcal{I} = 0$, given that γ^* is chosen, follows once again from the fact that the optimal allocation under $\mathcal{I} = 1$, which is achieved by setting $\gamma^* = 1$, is feasible with $\mathcal{I} = 0$, but generally not optimal.

■

Optimal Bank Financing with Less Opaque Assets[1]

Bank loan portfolios strongly shifted from business loans to mortgages in the last decades. The figure shows the share of mortgages in aggregated bank loan portfolios of 13 advanced economies. the share of mortgages roughly doubled from 32% to 62% while the share of business loans decreased from 65% to 38% since 1970.[2]



As mortgages seem to be easier to value than business loans, one implication from this shift is that bank assets became less opaque.[3] While real estate values are relatively accessible it is much harder for a bank to observe the factors which determine the value of a small (and unlisted) firm, for example the human capital and the effort of an entrepreneur. This relative transparency of mortgages is e.g. apparent in the fact that for mortgages there is platform lending and a secondary market (both is not the case for business loans) and mortgage contracts are standardized and short while business loan contracts are much more detailed and specific. This hints at lower monitoring costs.[4].

Despite this shift towards non-opaque assets the theoretical banking literature mainly builds on the premise that banks are the providers of finance for opaque (small) businesses and many theoretical

[1] Co-Authored with Kumar Rishabh, University of Basel.

[2] Calculations based on data by Jordà et al. [2016]. Loan data of the individual countries were converted into USD and then aggregated.

[3] This is even more true if we take into account that the other main categories of bank assets, securities and cash, can be valued directly by their market price.

[4] A widely used metric for the opaqueness of loans is the distance between bank and borrower. Recent studies for the US found the average distance for mortgages to be 2.5 times the average distance for small business loans, i.e. 90 miles vs. 241.8 miles, see Granja et al. [2018] and Eichholtz et al. [2019]

results seem to rely on the "opaqueness" of bank loans (or of bank assets more generally). An example is Calomiris and Kahn [1991] (CK in the following).[5] This paper provides an explanation why banks use demandable liabilities as means of financing. CK argue that demandable liabilities are useful to finance opaque assets because these contracts have a "disciplining role". Banks have the possibility to misuse the funds entrusted to them in their model and demandable liabilities give the investors the option to withdraw their funds if they fear this might happen. Thus demandable contracts mitigate the agency problem and allow for socially beneficial investment which would otherwise be impossible. But using demandable liabilities also has a cost, because it sometimes involves a liquidation of the bank and the investments.

In our paper we revisit the argument of CK introducing less opaque assets into their environment, i.e. assets who's value can be better observed by the investors (in CK the value of the bank assets is unobservable by the investors, i.e. these assets are fully opaque). As we show demandable liabilities are not optimal to finance more transparent assets although the agency problem applies to these assets as well. But if investors are able to better observe the value of the asset they can use this information to design non-demandable contracts, which allows to mitigate the agency problem without the costs of liquidation. Thus giving investors the option to withdraw their funds if they fear the bank will misuse them is not necessary for more transparent assets.

This theory suggests that opaque assets like business loans should be financed with demandable liabilities (like demand deposits) while less opaque assets like mortgages, securities or cash should be financed with non-demandable liabilities (like non-demandable debt or equity). In a second step we thus look at US balance sheet data of individual banks to see whether there is a correlation between the share of opaque assets, which we quantify as the share of business loans, and the share of demandable liabilities. As our measure of demandable liabilities we use a variable called *volatile liabilities* provided by the FDIC which essentially contains demand deposits not covered by deposit insurance and other very short-term liabilities.[6] We document a small but positive correlation between opaque assets (business loans) and volatile liabilities for small and medium sized banks (up to the 75th percentile) but interestingly no correlation for lager banks. This is consistent with the interpretation that the disciplining role of demandable liabilities is more important for smaller banks and big banks might enjoy insurance beyond deposit insurance (e.g. in the form of implicit too-big-to-fail guarantees) which reduce investor incentives to monitor and

[5]Other examples include Diamond [1984], Holmstrom and Tirole [1997] or Dang et al. [2017]

[6]The use of total demand deposits is complicated by deposit insurance which largely eliminates the incentives for investors to monitor the bank and thus also the need for contracts with a disciplining role. However, investors holding uninsured deposits and other very short term liabilities should have better incentives to monitor and discipline the issuing bank.

"discipline" the banks even for deposits not covered by deposit insurance.

The rest of the paper is organized as follows: In section 4.1 we introduce the theoretical model and derive the optimal contracts for assets with different degrees of opacity. In section 4.2 we relate our theoretical findings to US data.

4.1 The model

4.1.1 Environment

There is a bank and an investor and three periods $(0, 1, 2)$. Both agents are risk-neutral. The bank has no endowment but a project which takes 1 good as input in period 0 and returns $\tilde{y}$ in period 2. This return is stochastic and is a weighted average of a publicly observable aggregate factor $\tilde{y}_A$ and an idiosyncratic factor $\tilde{y}_I$, the realization of which only the bank can observe.

$$\tilde{y} = \alpha\tilde{y}_A + (1-\alpha)\tilde{y}_I \tag{4.1}$$

The parameter α steers the weight of the aggregate factor and will be our measure of the transparency (opaqueness) of the asset. The higher α the more important the aggregate factor for the return and the more transparent (less opaque) the asset. We interpret opaque assets with low α, where the idiosyncratic factor is dominant, as business loans and more transparent assets with high α, where the aggregate factor is dominant, as mortgages or securities. If α is 0 the environment is identical to the basic model of CK.

The realizations of the two factors are independent from each other and for simplicity we assume both factors have the same two-point distribution with a high realization y_h (with probability q) and a low realization y_l with probability $1-q$. This also implies that all assets have the same expected return, independent of their opaqueness. This set-up yields four possible states for $\tilde{y}$ (where $\Delta y = y_h - y_l$):

$$\begin{aligned}
y_1 &= \alpha y_h + (1-\alpha)y_h = y_h && \text{with probability} \quad q^2 \\
y_2 &= \alpha y_l + (1-\alpha)y_h = y_h - \alpha\Delta y && \text{with probability} \quad (1-q)q \\
y_3 &= \alpha y_h + (1-\alpha)y_l = y_l + \alpha\Delta y && \text{with probability} \quad q(1-q) \\
y_4 &= \alpha y_l + (1-\alpha)y_l = y_l && \text{with probability} \quad (1-q)^2
\end{aligned}$$

The realizations in states 1 and 4 do not depend on the opaqueness of the asset α since both factors yield the same high or low outcome (either y_l or y_h). In state 2 the return decreases with α as the aggregate factor is low and the idiosyncratic factor is high and the with α the weight of the low aggregate factor increases. Thus the return in this state decreases from y_h to y_l with α. The opposite is true for state 3 where the return increases with α from y_l to y_h.

In period 0 the bank promises to pay z goods to the investor in period 2. Since the realization of the aggregate factor is observable by both agents, z can be made contingent on $\tilde{y}_A$, which we denote as $z(\tilde{y}_A)$. When $\alpha = 0$, there is no aggregate factor and thus the payment is just z. As in CK, we assume there is an agency problem between the bank and the investor in the sense that the bank has the possibility to misuse the funds entrusted to her. In period 2 after observing the state of $\tilde{y}$ the bank can also run away (abscond) with the total return $\tilde{y}$ diminished by a proportion $A \in (0, 1)$ instead of repaying z. In this case the investor gets zero. Thus absconding in state $i \in \{1, 2, 3, 4\}$ implies a welfare loss of Ay_i.[7]

The investor is endowed with 1 good in period 0 which she can either store or invest in the bank. Thus the expected payments from the bank need to be at least 1 for the investor to participate. We will assume investment is beneficial, i.e.

$$E[\tilde{y}] > 1 \tag{4.2}$$

A method to prevent absconding is useful in this environment. For this purpose we introduce the possibility to liquidate the investment in the middle period 1. We assume that if the project (the bank) is liquidated in period 1 the investor gets a payment r for sure and the bank gets zero. We will assume

$$r < \min\{1, y_l\} \tag{4.3}$$

$r < 1$ implies that liquidation yields a lower return than storage and it is socially not efficient to always liquidate. $r < y_l$ implies that liquidation is wasteful in any state, i.e. even in state 4 where $y_4 = y_l$. Liquidation always yields less than if the investment would have been completed.

To allow for contracts where the investor has the option of demanding liquidation in the middle period we assume investors get a signal on the realization of the idiosyncratic state in the middle period 1. The signal is private information for the investor so contracts contingent on the signal are

[7] With this formulation we do not link absconding to the opaqueness of the asset. We could do this e.g. by assuming that the bank can only abscond with a fraction $1 - A$ of the idiosyncratic factor $(1 - \alpha)(1 - A)\tilde{y}_I$, and the investor can always capture the publicly observable aggregate factor $\alpha\tilde{y}_A$. This would give transparent assets with higher α a direct advantage against the absconding problem.

not possible.[8] The signal structure is as follows: in period 1 the investor gets the signal "good" with probability π and the signal "bad" with probability $1-\pi$. Given the good signal the probability of the good idiosyncratic state is $p_g = Pr(\tilde{y}_I = y_h|\pi)$ and given the bad signal the probability of the bad idiosyncratic state is $p_b = Pr(\tilde{y}_I = y_l|1-\pi)$.

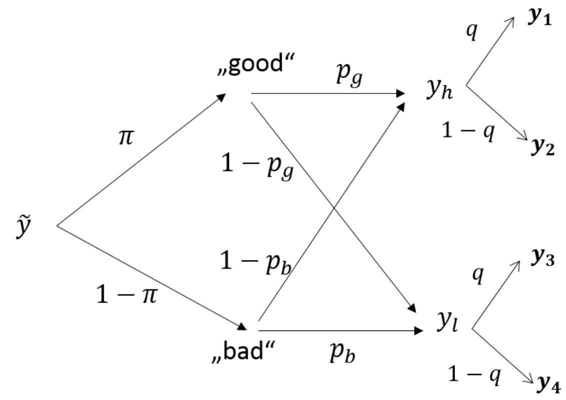


Figure 4.1: Signal structure for the idiosyncratic part of the return

p_g and p_b can be interpreted as the precision of the signal. To be informative they must be higher than the ex-ante probabilities q and $1-q$. Thus we need $p_g > q$ and $p_b > 1-q$. Also the ex-ante and the ex-post probabilities need to ad up as follows:

$$q = \pi p_g + (1-\pi)(1-p_b) \tag{4.4}$$

We can summarize the model as follows: In period 0 the bank offers the investor a payment of $z(\tilde{y}_A)$ goods in period 2 and (possibly) the option to withdraw r goods in period 1 but nothing in period 2. The investor then invests in the bank or stores the good. In period 1 the investor gets a signal about the idiosyncratic factor of the investment. Based on this signal she decides whether she should withdraw and get r if the contract allows this. If the investor withdraws the project (and the bank) is liquidated. If the project is not liquidated, in period 2 the project return $\tilde{y}$ realizes and the bank decides whether to abscond or pay back the investor. We call contracts without the option to withdraw in period 1 "non-liquidation contracts" and interpret them as non-demandable liabilities like long-term debt or equity (or a mixture of the two). We call contracts *with* the option to withdraw in period 1 "liquidation contracts" and interpret them as demandable liabilities like demand deposits.

[8] In this respect we differ from CK who assume payments can be contingent on the signal although the signal is private information. CK also assume that the investor has to pay a cost c to get a signal. We abstract from this and set $c = 0$. The cost of the signal can also be captured by its precision.

4.1.2 The case for liquidation contracts with fully opaque assets ($\alpha = 0$)

In this section we look at fully opaque assets where $\alpha = 0$, i.e. assets only driven by the idiosyncratic factor. The goal is to derive conditions when liquidation contracts are optimal for fully opaque assets.[9] We will use these conditions also in the later sections with less opaque assets when $\alpha > 0$. Note that when $\alpha = 0$, the bank just proposes a non-contingent payment z.

First we refine assumption (4.2). We want to assume that investment is beneficial even if the bank absconds in the bad idiosyncratic state. Thus if non-liquidation contracts are not feasible there is no investment although it would be efficient and liquidation contracts are potentially welfare improving. We assume the expected value of the investment minus the expected welfare loss due to absconding in the bad state must be above the return from storage:

$$E[\tilde{y}] - (1-q)Ay_l > 1 \tag{4.5}$$

Second we want to assume that absconding is sufficiently attractive for the bank so contracts without liquidation are not feasible. This implies that $1 - A$, the fraction the bank can abscond with, needs to be sufficiently high or A sufficiently low. Specifically we assume:

$$Ay_l < 1 \tag{4.6}$$

$$qAy_h < 1 \tag{4.7}$$

The first assumption rules out a non-liquidation contract where the bank never absconds. For such a contract bank profits without absconding, $\tilde{y} - z$, must always exceed the return in case of absconding, $(1 - A)\tilde{y}$. This is satisfied in both states if $z \leq Ay_l$. If the bank never absconds the payments to the investor are just z. Thus for such a contract to be feasible we need $Ay_l \geq 1$ which is ruled out by assumption (4.6). The second assumption rules out a non-liquidation contract where the bank absconds in the bad state. In this case the bank payment needs to satisfy: $Ay_l < z \leq Ay_h$. The investor is now only paid in the good state and thus expected payments are qz. This contract is feasible if $qAy_h \geq 1$ and thus ruled out by assumption (4.7).

If (4.6) and (4.7) hold, non-liquidation contracts are not feasible. We now turn to liquidation contracts. Suppose z is like in the second contract, $Ay_l < z \leq Ay_h$, so the bank absconds in the

[9] We call a contract *optimal* if it is feasible and the expected welfare (output) under this contract is higher than that of any other feasible contract.

bad state but not in the good but now the investor has the option to withdraw r in period 1.[10] Suppose the investor demands liquidation after the bad signal. The expected payments to the investor are:

$$\pi p_g z + \pi(1-p_g)0 + (1-\pi)r \geq 1 \tag{4.8}$$

With probability π the investor gets the good signal and does not liquidate. Then with probability p_g she actually ends up in the good state and gets payment z. With the inverse probability $1-p_g$ she ends up in the bad state where the bank absconds and the investor gets zero. With probability $1-\pi$ the investor receives the bad signal and liquidates the bank getting r. Since $r < 1$ we need $p_g z > 1$ for this contract to be feasible. And since $z \leq Ay_h$ this implies

$$p_g A y_h > 1 \tag{4.9}$$

Condition (4.8) implies this contract is feasible if π is above a certain threshold

$$\pi \geq \frac{1-r}{p_g A y_h - r} \tag{4.10}$$

Condition (4.10) can be interpreted as a threshold on signal precision p_b and on the return in case of liquidation, r. Since π increases with p_b from (4.4) and since p_b cannot exceed 1 we can reformulate (4.10) as two conditions on signal precision and r:

$$p_b \geq \frac{(1-q)(p_g A y_h - r) - (1-r)(1-p_g)}{p_g A y_h - 1} \tag{4.11}$$

$$r \geq \frac{p_g(1-qAy_h)}{p_g - q} \tag{4.12}$$

This leads to the following proposition:

Proposition 4.1. *Suppose absconding is sufficiently attractive for the bank such that non-liquidation contracts are not feasible, i.e.* (4.6) *and* (4.7) *hold. Also suppose* (4.10), (4.11) *and* (4.12) *hold and the signals are sufficiently precise. Then the optimal contract for fully opaque assets is a liquidation contract.*

[10] CK also consider a contract where the bank is always liquidated. As we assume $r < 1$ such a contract can never be viable and we ignore it here.

4.1.3 Optimal contracts with less opaque assets ($\alpha > 0$)

In the last section we showed that it is optimal to finance fully opaque assets with liquidation contracts if absconding is sufficiently attractive for the bank and the signals are sufficiently precise. In the following we study how this conclusion changes (under the same assumptions) when we increase the transparency of assets (α). Since the analysis is a little bit more involved we divide it into non-liquidation and liquidation contracts.

Non-liquidation contracts

A main difference to the case of fully opaque assets is that with $\alpha > 0$ banks and investors can sign contracts contingent on the aggregate state. We will denote the payments given the high (low) aggregate state as $z(y_h)$ ($z(y_l)$). Payment $z(y_h)$ applies to states 1 and 3 (which both have a high realization of the aggregate factor) and payment $z(y_l)$ to states 2 and 4 (which both have a low realization aggregate factor). Similar to before the bank does not abscond in state i if the payment is below a fraction A of the return in a given state:

$$z(\tilde{y}_A) \leq Ay_i = \bar{z}_i \tag{4.13}$$

We denote the maximal feasible payment such that the bank does not abscond in state i as $\bar{z}_i$. Because any feasible contract must at least yield 1 in expected terms for the investor, it is crucial how much bank can maximally pay. We will look at the maximal expected payments to the investor in 4 different situations. They are denoted with S_i where i describes the state(s) in which the bank absconds and $i = 0$ means the bank never absconds. $\bar{z}(S_i)$ is then the maximal expected payment for the investor in situation S_i. We look at the following four situations: S_0, where the bank never absconds, S_4, where the bank absconds in state 4, S_3, where the bank absconds in state 3 and $S_{3,4}$ where the bank absconds in states 4 and 3.[11]

Before considering the feasibility and optimality of non-liquidation contracts we look at expected welfare in these four situations denoted as $W(S_i)$. Without liquidation welfare losses only result from absconding. Thus in the situation where the bank never absconds (S_0) there are no welfare losses. In the other situations the expected welfare losses are $(1-q)^2Ay_l$ in S_4, $q(1-q)Ay_3$ in situation S_3 and $(1-q)A\big((1-q)y_l + qy_3\big)$ in $S_{3,4}$. It is clear that expected welfare losses are lowest in S_0 and highest in $S_{3,4}$. But it is not clear whether they are higher in situations S_4 or S_3. To put

[11] We don't need to consider absconding in states 2 and 4 for example. The maximal expected payment of such a situation would be strictly dominated by S_4.

more structure on the problem we will assume that the welfare losses in situation S_4 are strictly lower than in situation S_3, independent of α. This implies the following assumption on q:

$$q > 0.5 \tag{4.14}$$

The intuition is that if $q > 0.5$ the probability of state 4, $(1-q)^2$, where the bank absconds in situation S_4 is strictly lower than the probability of state 3, $(q(1-q))$, where the bank absconds in situation S_3. Since the welfare losses in state 4 are lower than in state 3 this implies $W(S_4)$ must be higher than $W(S_3)$. We thus arrive at the following ordering of the four situations in terms of welfare: $W(S_0) > W(S_4) > W(S_3) > W(S_{3,4})$. We will consider the feasibility of the contracts in this order.

situation S_0: the bank never absconds

In this situation the maximum the bank can pay is $z(y_l) = \bar{z}_4$ and $z(y_h) = \bar{z}_3$. Thus the maximal expected payments to the investor are

$$\bar{z}(S_0) = qA(y_l + \alpha\Delta y) + (1-q)Ay_l = Ay_l + \alpha A q \Delta y \tag{4.15}$$

Note that $\bar{z}(S_0)$ increases in α. If $AE[\tilde{y}] > 1$, $\bar{z}(S_0) \geq 1$ exists for an $\alpha \in (\alpha(S_0), 1)$ where

$$\alpha(S_0) = \frac{1 - Ay_l}{Aq\Delta y} \tag{4.16}$$

situation S_4: the bank absconds in state 4

The maximum the bank can pay is $z(y_l) = \bar{z}_2$ and $z(y_h) = \bar{z}_3$. Thus the maximal expected payments to the investor are

$$\bar{z}(S_4) = qA(y_l + \alpha\Delta y) + (1-q)qA(y_h - \alpha\Delta y) = A(qy_l + q(1-q)y_h) + \alpha A q^2 \Delta y \tag{4.17}$$

As in situation S_0 maximal expected payments increase with the transparency of the asset. The equivalent threshold to $\alpha(S_0)$ is given by

$$\alpha(S_4) = \frac{1 - qA(y_l + (1-q)y_h)}{q^2 A \Delta y} \tag{4.18}$$

Comparing (4.18) with (4.16) we get that $\alpha(S_0) < \alpha(S_4)$ if $qAy_h < 1$ which holds by condition (4.5). Thus S_0 is feasible where S_4 is feasible. But since $W(S_0) > W(S_4)$ the non-liquidation contract S_4 is never optimal.

situation S_3: the bank absconds in state 3

In this situation the maximum the bank can pay is $z(y_l) = \bar{z}_4$ and $z(y_h) = \bar{z}_1$. Thus the maximal expected payments to the investor are

$$\bar{z}(S_3) = q^2 Ay_h + (1-q)Ay_l \tag{4.19}$$

Note that $\bar{z}(S_3)$ is independent of α. Also note that $\bar{z}(S_3)$ is a weighted average of qAy_h and Ay_l which are both below 1 by assumptions (4.6) and (4.7). Thus $\bar{z}(S_3)$ will also be below 1 and is never feasible.

situation $S_{3,4}$: the bank absconds in states 3 and 4

The maximum the bank can pay is $z(y_l) = \bar{z}_2$ and $z(y_h) = \bar{z}_1$. Thus the maximal expected payments to the investor are

$$\bar{z}(S_3) = q^2 Ay_h + (1-q)qA(y_h - \alpha\Delta q) \tag{4.20}$$

$\bar{z}(S_3)$ decreases in α thus it is maximal at $\alpha = 0$ where $\bar{z}(S_3)|_{\alpha=0} = qAy_h$. This is exactly the second non-liquidating contract we considered in the last section ruled out by assumption (4.7).

We conclude this discussion with the following proposition:

Proposition 4.2. *If $AE[\tilde{y}] > 1$ there exists a non-liquidation contract S_0 for sufficiently transparent assets with $\alpha \in (\alpha(S_0), 1)$ where the bank never absconds and which implements first best.*

Why are non-liquidation contracts without absconding feasible for sufficiently transparent assets but not for opaque ones? Consider an example: Suppose $y_h = 2, y_l = 1, q = 0.6$ and $A = 0.8$. Now suppose we have an opaque asset where $\alpha = 0.1$. The states this asset can take are: $y_1 = 2, y_2 = 1.9, y_3 = 1.1, y_4 = 1$. Since the idiosyncratic factor is dominant for this asset the states where the idiosyncratic factor has high realization (i.e. corresponding to y_1 and y_2) and the states where the realization is low (corresponding to y_3 and y_4) are closely together. This also means payments contingent on the aggregate factor are not very useful. With $\alpha = 0.1$ the maximal payments such that the bank does not abscond are $z(y_l) = \bar{z}_4 = 0.8$ for states 4 and 2 and $z(y_h) = \bar{z}_3 = 0.88$ for states 3 and 1. Thus $\bar{z}(S_0) = 0.848$ which is below 1 and there is no investment. Now consider a very transparent asset where $\alpha = 0.9$. The states this asset can take are: $y_1 = 2, y_2 = 1.1, y_3 = 1.9, y_4 = 1$. Now the states where the aggregate factor is the same (y_1, y_3 and y_2, y_4) are close together and payments contingent on the aggregate state are much more effective: while the payment contingent on the low realization of the aggregate factor is still $z(y_l) = \bar{z}_4 = 0.8$, the payment contingent on the high aggregate factor is $z(y_h) = \bar{z}_3 = 1.52$. Thus maximal expected payments to the investor increase to 1.232 and investment would be feasible.

The general message is that while payments contingent on the aggregate factor are possible with very opaque and transparent assets, they are much more useful in the latter case. For very opaque assets it would be much more effective to tie the payments to the realization of the idiosyncratic factor. But this is not possible because it is only privately observable. The better observability of the returns for very transparent assets allows to mitigate the absconding problem by better using the flexibility of contingent contracts.

Liquidation contracts

We consider contracts with liquidation in the middle period after the bad signal for the same 4 situation as in the section before. We denote the situations as $S_{i,L}$ where "L" stands for "liquidation". Liquidation decreases the expected welfare losses due to absconding (because it prevents them sometimes) but it also introduces additional welfare losses. The expected liquidation costs are the difference of the expected return without liquidation and r. In the case where the investor liquidates after the bad signal (which happens with probability $1-\pi$) they are given by:

$$(1-\pi)\left[(1-p_b)(qy_h + (1-q)y_2) + p_b(qy_3 + (1-q)y_l) - r\right] \tag{4.21}$$

With probability $1-p_b$ states 1 or 2 occur and the expected losses are $qy_h + (1-q)y_2 - r$. With probability p_b states 3 or 4 occur and expected losses are $qy_3 + (1-q)y_l - r$. Note that the liquidation costs do not depend on the situations (S_i). This implies the ordering of the contracts in terms of expected welfare from the last section is preserved with liquidation contracts: $W(S_0) > W(S_{0,L}) > W(S_{4,L}) > W(S_{3,L}) > W(S_{3,4,L})$. The non-liquidation contract S_0 dominates all liquidation contracts and these are ordered according to the welfare losses from absconding. In the following we consider the liquidation contracts in this order and compare them with the non-liquidation contract S_0.

situation $S_{0,L}$: the bank never absconds

The bank behaves as in the non-liquidation contract S_0. She pays $z(y_l) = \bar{z}_4$ and $z(y_h) = \bar{z}_3$ and thus never absconds. But the investor still liquidates the bank after the bad signal. If $r > Ay_l$ (which is feasible from our assumptions on r) liquidation can still increase expected payments at least in state 4. Maximal expected payments are given by:

$$\bar{z}(S_{0,L}) = \pi\big(p_g(q\bar{z}_3 + (1-q)\bar{z}_4)) + (1-p_g)(q\bar{z}_3 + (1-q)\bar{z}_4))\big) + (1-\pi)r \tag{4.22}$$

(4.22) can be written as a weighted average between the maximal exp. payments without liquidation and r: $\bar{z}(S_{0,L}) = \pi\bar{z}(S_0) + (1-\pi)r$. Both $\bar{z}(S_0)$ and $\bar{z}(S_{0,L})$ are below 1 at $\alpha = 0$ and increase in α. From the weighted average formulation of $\bar{z}(S_{0,L})$ however, we can conclude that the slope of $\bar{z}(S_{0,L})$ must be lower and therefore $\alpha(S_{0,L}) > \alpha(S_0)$. Ergo the non-liquidation contract S_0 is always feasible where the liquidation contract $S_{0,L}$ is feasible and since S_0 has no welfare losses the contract $S_{0,L}$ is never optimal.

situation $S_{4,L}$: the bank absconds in state 4

In this contract the bank can maximally pay $z(y_l) = \bar{z}_2$ and $z(y_h) = \bar{z}_3$. Thus maximal payments are:

$$\begin{aligned}\bar{z}(S_{4,L}) &= \pi p_g\big(q\bar{z}_3 + (1-q)\bar{z}_2 + (1-p_g)q\bar{z}_3\big) + (1-\pi)r \\ &= \pi\big(qAy_l + p_g(1-q)Ay_h + \alpha A\Delta y(q - p_g(1-q))\big) + (1-\pi)r\end{aligned} \tag{4.23}$$

From assumption (4.14) $\bar{z}(S_{4,L})$ increases in α. As shown in the appendix the feasibility of this contract depends on two thresholds on π, $\bar{\pi}(S_{4,L}) > \underline{\pi}(S_{4,L})$ in the following way.[12]

- If $\pi > \bar{\pi}(S_{4,L})$ contract $S_{4,L}$ is feasible for any α
- If $\underline{\pi}(S_{4,L}) < \pi < \bar{\pi}(S_{4,L})$ contract $S_{4,L}$ is feasible for $\alpha \geq \alpha(S_{4,L})$ where

$$\alpha(S_{4,L}) = \frac{1 - (1-\pi)r - \pi A(qy_l + (1-q)p_g y_h)}{\pi\Delta y A(q - p_g(1-q))} \tag{4.24}$$

- If $\pi < \underline{\pi}(S_{4,L})$ contract $S_{4,L}$ is never feasible

We know expected welfare of $S_{4,L}$ is higher than the other liquidation contracts (ignoring $S_{0,L}$). Thus it is clear that in the first region $S_{4,L}$ is the optimal contract for $\alpha \in (0, \alpha(S_0))$ where the non-liquidation contract S_0 is not feasible. In the middle region $S_{4,L}$ is only optimal if it is feasible where S_0 is not feasible, i.e. if $\alpha(S_{4,L}) < \alpha(S_0)$. It can be demonstrated that this is satisfied if $\pi > \pi^*(S_{4,L})$ where $\underline{\pi}(S_{4,L}) < \pi^*(S_{4,L}) < \bar{\pi}(S_{4,L})$. Thus the liquidation contract $S_{4,L}$ is optimal in the middle region for $\alpha \in (\alpha(S_{4,L}), \alpha(S_0))$ as long as $\pi \geq \pi^*(S_{4,L})$.

[12] As shown in the section on liquidation contracts with fully opaque assets we can interpret the conditions on π as conditions on signal precision and on the liquidation value of the investment r, see (4.11) and (4.12). If π needs to have a certain value this means signal precision and the liquidation value need to be sufficiently high.

situation $S_{3,L}$: the bank absconds in state 3

In this situation the maximum the bank can pay is $z(y_l) = \bar{z}_4$ and $z(y_h) = \bar{z}_1$ after the good signal and it gets liquidated after the bad signal.

$$\begin{aligned} \bar{z}(S_{3,L}) &= \pi\big(p_g(q\bar{z}_1 + (1-q)\bar{z}_4) + (1-p_g)(1-q)\bar{z}_4\big) + (1-\pi)r \\ &= \pi\big((1-q)Ay_l + qp_gAy_h\big) + (1-\pi)r \end{aligned} \tag{4.25}$$

As the non-liquidation contract S_3 the liquidation contract is independent of α. It is feasible if $\pi \geq \pi(S_{3,L})$ where

$$\pi(S_{3,L}) = \frac{1-r}{A((1-q)y_l + p_gqy_h) - r} \tag{4.26}$$

situation $S_{3,4,L}$: the bank absconds in states 3 and 4

The maximum the bank can pay is $z(y_l) = \bar{z}_2$ and $z(y_h) = \bar{z}_1$. Thus maximal payments are:

$$\begin{aligned} \bar{z}(S_{3,4,L}) &= \pi p_g(q\bar{z}_1 + (1-q)\bar{z}_2) + (1-\pi)r \\ &= \pi p_g A(y_h - \alpha\Delta y(1-q)) + (1-\pi)r \end{aligned} \tag{4.27}$$

As with the non-liquidation contract $S_{3,4}$ maximal expected payments $\bar{z}(S_{3,4,L})$ decrease in α. Thus they are maximal at $\alpha = 0$ and the contract is feasible at $\alpha = 0$ if conditions (4.10), (4.11) and (4.12) hold, in other words if $\pi \geq \pi(S_{3,4,L})$ where $\pi(S_{3,4,L})$ is given by the right-hand side of (4.10).

$$\pi(S_{3,4,L}) = \frac{1-r}{p_gAy_h - r} \tag{4.28}$$

The contract is then feasible as long as $\alpha \in (0, \alpha(S_{3,4,L}))$ where

$$\alpha(S_{3,4,L}) = \frac{\pi p_g Ay_h + (1-\pi)r - 1}{\pi p_g(1-q)A\Delta y} \tag{4.29}$$

The following lemma summarizes the relationships between the thresholds on π of the liquidation contracts (ignoring $S_{0,L}$):

Lemma 1. *$\bar{\pi}(S_{4,L}) > \pi(S_{3,L}) > \underline{\pi}(S_{4,L})$ and $\pi(S_{3,L}) > \pi(S_{3,4,L})$ and $\pi^*(S_{4,L}) > \pi(S_{3,L})$ given $p_g > p_g^*$ where*

$$p_g^* = \frac{1 - 2(1-q)Ay_l}{(1-q)(1-Ay_l) + qAy_h(2q-1)}\, q \tag{4.30}$$

Combining lemma 1 with the previous propositions 4.1 and 4.2 we can characterize the optimal contracts for the full model:

Proposition 4.3. *Suppose the assumptions from proposition 4.1 and 4.2 hold so the optimal contract for fully opaque assets is a liquidation contract and for fully transparent assets it is a non-liquidation contract. Also suppose the precision of the good signal is sufficiently high such that $p_g > p_g^*$. Then for sufficiently transparent assets with $\alpha \in (\alpha(S_0), 1)$ the non-liquidation contract S_0 is optimal and the optimal liquidation contracts depend on signal precision and r according to the following four cases:*

1. *If the signal precision and r are high ($\pi > \bar{\pi}(S_{4,L})$) the liquidation contract $S_{4,L}$ is optimal for $\alpha \in (0, \alpha(S_0))$.*

2. *If the signal precision and r are in a medium range ($\pi^*(S_{4,L}) \leq \pi < \bar{\pi}(S_{4,L})$) liquidation contract $S_{3,L}$ is optimal for $\alpha \in (0, \alpha(S_{4,L}))$ and liquidation contract $S_{4,L}$ is optimal for $\alpha \in (\alpha(S_{4,L}), \alpha(S_0))$.*

3. *If the signal precision and r are low ($\pi(S_{3,L}) \leq \pi < \pi^*(S_{4,L})$)) the liquidation contract $S_{3,L}$ is optimal for $\alpha \in (0, \alpha(S_0))$).*

4. *If the signal precision and r are are very low ($\pi < \pi(S_{3,L})$) the liquidation contract $S_{3,4,L}$ is optimal for $\alpha \in (0, \alpha(S_{3,4,L}))$ if $\alpha(S_{3,4,L}) < \alpha(S_0)$ and for $\alpha \in (0, \alpha(S_0))$ if $\alpha(S_{3,4,L}) > \alpha(S_0)$.*

We illustrate proposition (4.3) with an example. Suppose $y_h = 2, y_l = 1, q = 0.6, A = 0.8$ and $r = 0.9$. Note that these values satisfy assumptions (4.5), (4.6), (4.7) and (4.14) and $\alpha(S_0) = 0.416$. According to proposition 4.3 for assets with a "transparency" of $\alpha > 0.416$ the optimal contract will always be the non-liquidation contract S_0. Now we look at the liquidation contracts. We will assume p_g is fixed (this will pin down all the thresholds for π) and then explore the four cases by gradually decreasing p_b. We will assume the precision after the good signal is high, $p_g = 0.95$ (this is necessary to explore all the cases in proposition 4.3). The thresholds are: $\bar{\pi}(S_{4,L}) = 0.532$, $\pi^*(S_{4,L}) = 0.383$, $\pi(S_{3,L}) = 0.301$ and $\pi(S_{3,4,L}) = 0.161$. We first choose $p_b = 0.95$ which implies $\pi = 0.611 > \bar{\pi}(S_{4,L})$. The following figure shows the optimal contracts for case 1 with maximal expected payments of the optimal contracts $\bar{z}(S_i)$ as a function of α. The green line shows the maximal expected payments of the non-liquidation contract $\bar{z}(S_0)$ and the blue line depicts the maximal expected payments of the liquidation contract $\bar{z}(S_{4,L})$. If the lines are above 1 the contracts are feasible. The liquidation contract $S_{4,L}$ is optimal until $\alpha(S_0) = 0.416$.

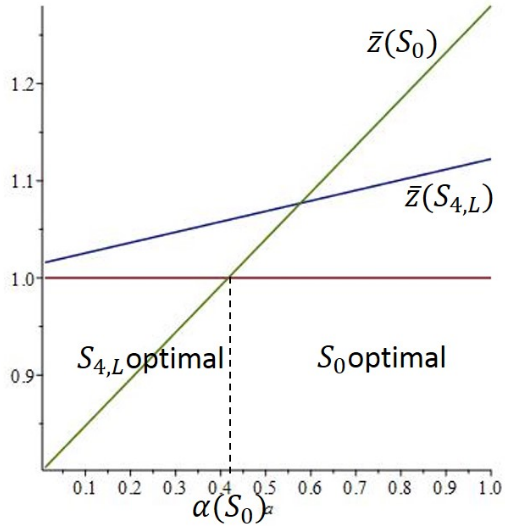


Figure 4.2: Case 1: high signal precision and r

Now we decrease p_b until we are in the second case and π is $\in (0.383, 0.52)$. We choose $p_b = 0.7$ which implies $\pi = 0.461$. These values yield $\alpha(S_{4,L}) = 0.163$. So the liquidation contract $S_{4,L}$ is now optimal for $\alpha \in (0.163, 0.416)$ and the liquidation contract $S_{3,L}$ is optimal for $\alpha \in (0, 0.163)$. As we know $\bar{z}(S_{3,L})$ which is shown as a red line is independent of α.

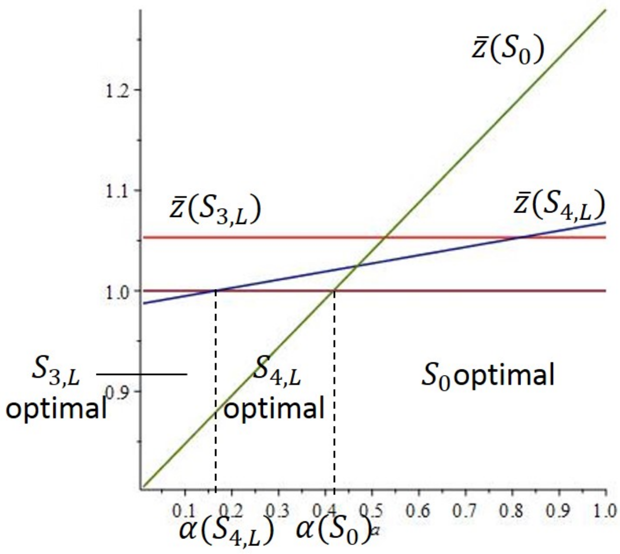


Figure 4.3: Case 2: medium signal precision and r

Finally, the following pictures show the situation for $p_b = 0.6$ which implies $\pi = 0.364$ (left side) and $p_b = 0.5$ which implies $\pi = 0.222$ (right side). On the left side the optimal liquidation contract is $S_{3,L}$ for $\alpha \in (0, \alpha(S_0))$. In this case liquidation contract $S_{4,L}$ (not shown) is not feasible anymore. On the right side precision is so low that liquidation contracts $S_{4,L}$ and $S_{3,L}$ (both not shown) are not feasible anymore. Only contract $S_{3,4,L}$ is feasible and thus optimal for $\alpha \in (0, \alpha(S_0))$. The black line shows the maximal expected payments of liquidation contract $S_{3,4,L}$, $\bar{z}(S_{3,4,L}$ which we know decreases in α.

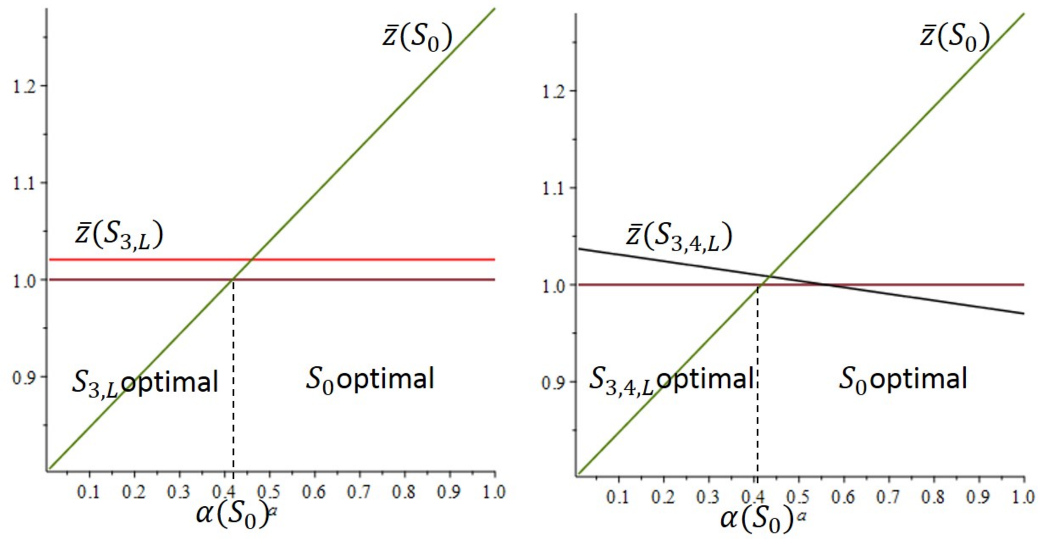


Figure 4.4: Cases 3 (left) and 4 (right)

What is the basic take-away of proposition 4.3 and this example? For opaque assets it is optimal to use a liquidation contract and for transparent assets it is optimal to use a non-liquidation contract. While the non-liquidation contract (S_0) dominates all liquidation contracts in terms of welfare it is not feasible for opaque assets. The reason is that the return of opaque assets is largely determined by the unobservable idiosyncratic factor and payments which can use the information of the aggregate factor are not very useful. For such assets only liquidation contracts can partially prevent absconding and are thus useful. For transparent assets, where the asset return is largely determined by the publicly observable aggregate factor, payments which can use this information can also prevent absconding. Since they do this without the liquidation costs of liquidation contracts (and in case of contract S_0 even without absconding costs) they are preferable for transparent assets.

4.2 Empirical analysis

In this section we try to link our theoretical considerations to the data. The theoretical model suggests that banks should finance opaque assets like business loans with demandable liabilities and less opaque assets like mortgages or securities with non-demandable liablities (which we could interpret as long-term debt or equity). However, the model we used was extremely stylized. It ignores other factors why banks might hold demandable liabilities (which could also be linked to the asset side) and it takes the asset side as given. In a more complete approach the bank would choose the asset side as part of a portfolio choice problem. Thus the conclusions we can draw from empirical correlations between opaque assets (business loans) and demandable liabilities are limited. Specifically we cannot interpret them as a test on the "disciplining theory" of demandable debt in the sense that if we don't observe a strong, or even one-to-one correlation between these variables the theory must be rejected. However, we still believe that if the disciplining role of demandable liabilities is economically important we should somehow see that banks with more opaque assets also issue more demandable liabilities.

We use yearly US bank (holding company) and savings and thrift institutions data from 1992 to 2018 provided by the call reports from the FDIC.[13] As our variable for opaque assets we use commercial and industrial (C&I) loans. As a measure of demandable liabilities we would ideally use uninsured deposits, i.e. demand deposits not covered by deposit insurance. The reason is that deposit insurance reduces (or even eliminates) the incentives for investors to monitor banks [Calomiris and Jaremski, 2019, Demirgüç-Kunt and Huizinga, 2004] and also makes deposits a cheaper source of funding [Admati and Hellwig, 2014]. Therefore, if there is a link between business loans and demand deposits because demand deposits have a disciplining role this link is most likely distorted by deposit insurance.[14] To address this we will use what the FDIC calls *volatile liabilities* as our measure of demandable liabilities. Volatile liabilities are essentially non-insured demand deposits plus some other very short-term liabilities (also see the appendix for a more detailed explanation). We first look at some aggregate statistics for the two variables. The following graphs show the evolution of business loans and volatile liabilities to total assets in the aggregate and for four bank sizes: small banks (below the 25th percentile), medium sized banks (between the 25th and the 75th percentile), big banks (between the 75th and the 99th percentile) and very

[13] In the appendix we provide a more comprehensive description of the data.

[14] CK also stress this point and emphasize that their model should capture a historical, pre-deposit-insurance economy. The current levels of bank deposits and business loans holdings make a link in terms of levels very implausible. The average share of deposits to total assets in the sample is roughly ten times the average share of business loans (84.2% vs. 8.9%).

big banks (above the 99th percentile). We use the last category because the US banking system has a highly skewed size distribution. For instance, *community banks*, that are defined as banks with assets less than USD 10 Billion,[15] account for about 97.5% of all the banks but only about 15% total banking sector assets. The top four banks – viz. JPMorgan Chase, Bank of America, Citigroup and Wells Fargo & Co.) account for about 44% of banking sector assets with an average asset of about USD 1.8 Trillion each.

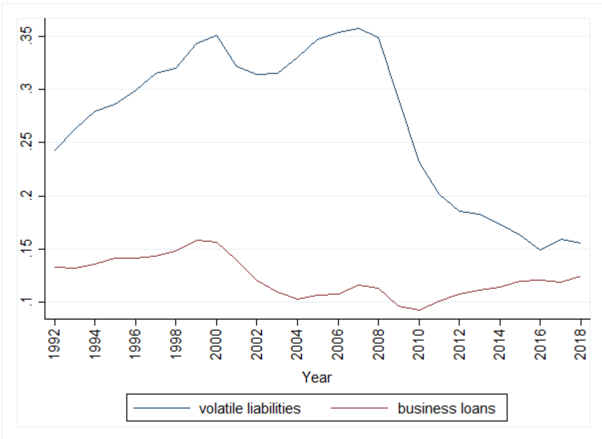


Figure 4.5: Aggregate evolution of the share of volatile liabilities and business loans to total assets

Over most of the sample period volatile liabilities are more than twice as high as business loans but after 2010 they decrease massively. This is most likely due to a change in the cap of deposit insurance which went up from 100′000 USD to 250′000 USD. This suddenly and strongly reduced the amount of uninsured deposits which are the essential part of volatile liabilities. In terms of variation there seems to be quite a close co-movement between volatile liabilities and business loans until 2010 after which the correlation is reversed (volatile liabilities go down and business loans go up).

[15]https://www.federalreserve.gov/supervisionreg/community-and-regional-financial-institutions.htm (accessed on June 18, 2020)

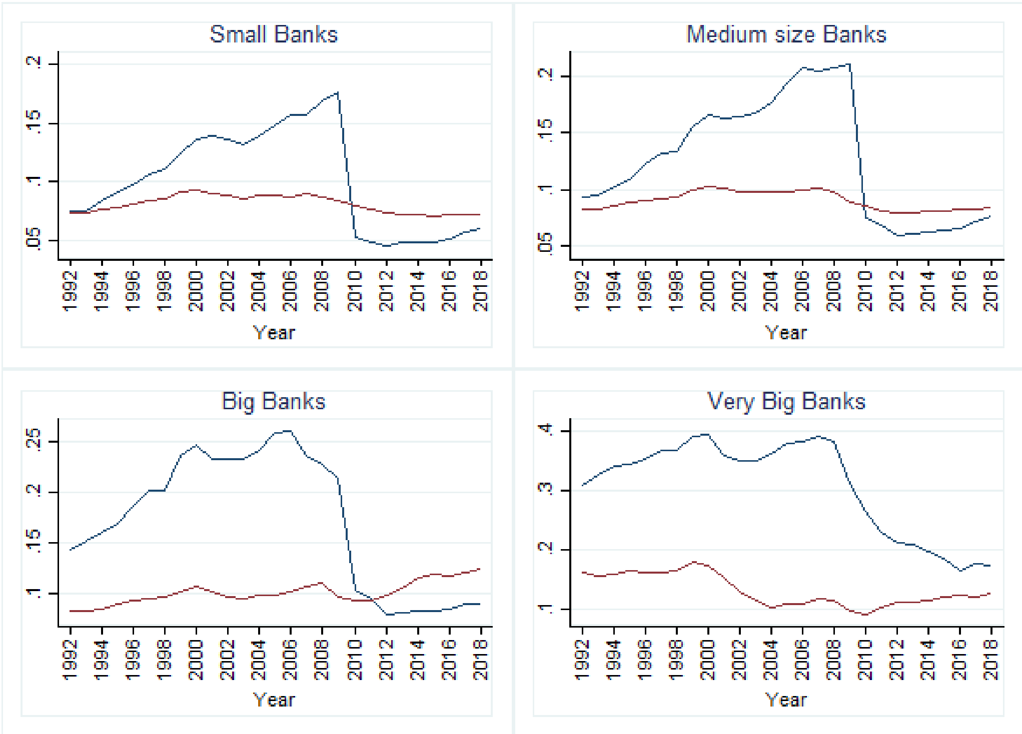


Figure 4.6: Evolution share of volatile liabilities (blue) and business loans (red): Size classes

For the four size groups the big drop of volatile liabilities in 2010 is apparent in all four graphs. The reversed correlation after 2010 seems to be especially pronounced for the very big banks but less so for the other three size groups. In terms of levels the divergence between volatile liabilities and business loans before 2010 seems to increase with size (note the different scale of the y-axis in the graphs). For small banks the two levels are much closer together than for the very big banks.

Now we look at the correlations between opaque assets and volatile liabilities more closely. To do this we regress the ratio of volatile liabilities to total assets on the ratio of business loans to total assets. With this specification a coefficient of one would correspond to a one-to-one relationship between the two variables. If also the intercept of this relationship is zero then even in levels business loans and volatile liabilities move together (i.e. the fitted values would lie on a 45-degree line going through the origin). We run this regression for the cross-section of banks in each year between 1992 and 2018. This should capture the strong drop in volatile liabilities in 2010. We first plot the estimated coefficient on the share of business loans in total assets and the intercept first without making any distinction across sizes (the dashed lines show the 95%-confidence bands).

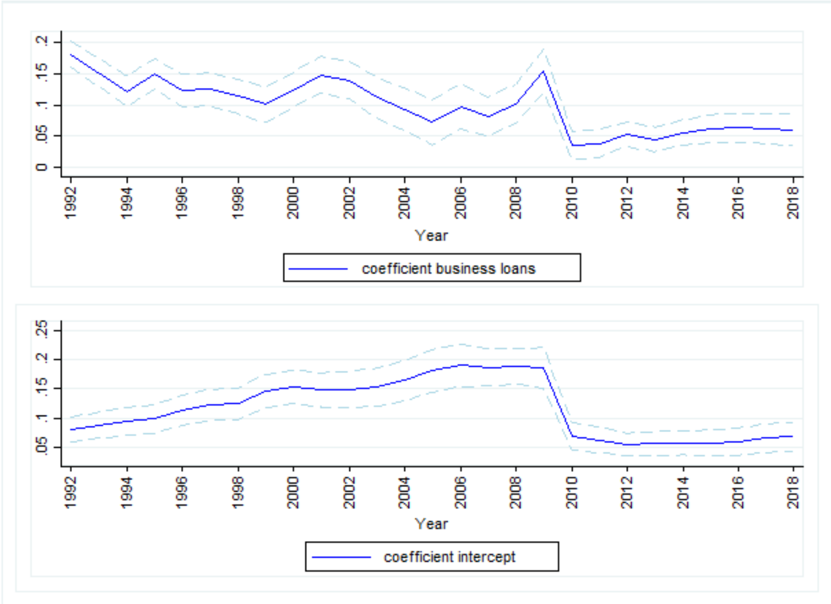


Figure 4.7: Regression coefficients: whole sample

We see the coefficient on business loans is positive for all years but quite small and certainly less than one. Between 1992 and 2008 it trends down from 0.18 to 0.10 and in 2010 the coefficient sharply drops to 0.035 so it seems the change in the cap of deposit insurance significantly lowered the correlation. The average coefficient on business loans before 2010 is 0.12 and after 2010 it is 0.05. On the other hand the intercept trends upwards from 0.08 in 1992 to 0.19 in 2009 (the average before 2010 is 0.14) and then sharply falls to a lower level of 0.07 in 2010 (the average after 2010 is 0.06). The intercept is close (but statistically different) from 0. What do these numbers tell us? Suppose we have two groups of banks, one group with a share of business loans of 0.1 and the other with 0.2 of total assets. What shares of volatile liabilities do the coefficients predict for these groups before and after 2010 on average? Before 2010 we take the average intercept (0.14) and ad the average coefficient on business loans (0.12) times the share of business loans. On average we would expect the banks with a share of 0.1 to have a ratio of volatile liabilities of 0.152 and the banks with a share of 0.2 to have a ratio of 0.164. As the coefficient on business loans is close to zero, these estimates are still close to the intercept of 0.14 although the difference in the share of business loans between the two groups is sizeable. But since the coefficient is so low, sizeable variation in terms of opaque assets does not translate into sizeable variation in terms of volatile liabilities. After 2010 the predicted shares of volatile liabilities are even closer to the intercept because the average coefficient on business loans (0.05) is very close to zero. We get a predicted share of volatile liabilities of 0.065 (0.07) for the banks with a share of 0.1 (0.2).

Now we plot the estimated coefficient on the share of business loans to total assets and the intercept for the different size classes. We restrict attention to the small and the big banks though. The results for the medium sized banks are very similar to the small banks and the aggregate estimates and the results for the very big banks are similar to the results of the big banks. Also, the estimates for the very big banks are not very precise because we don't have so many observations. We put those estimates in the appendix.

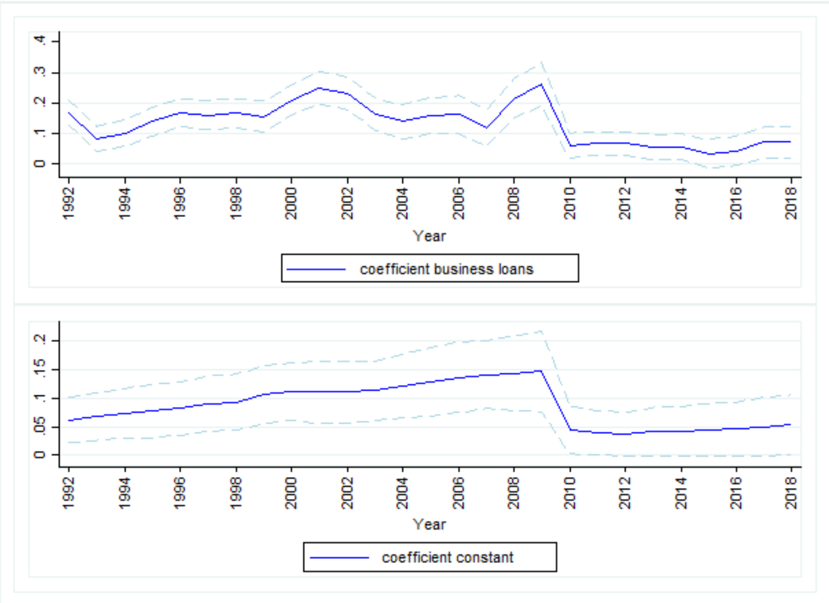


Figure 4.8: Regression coefficients: small banks

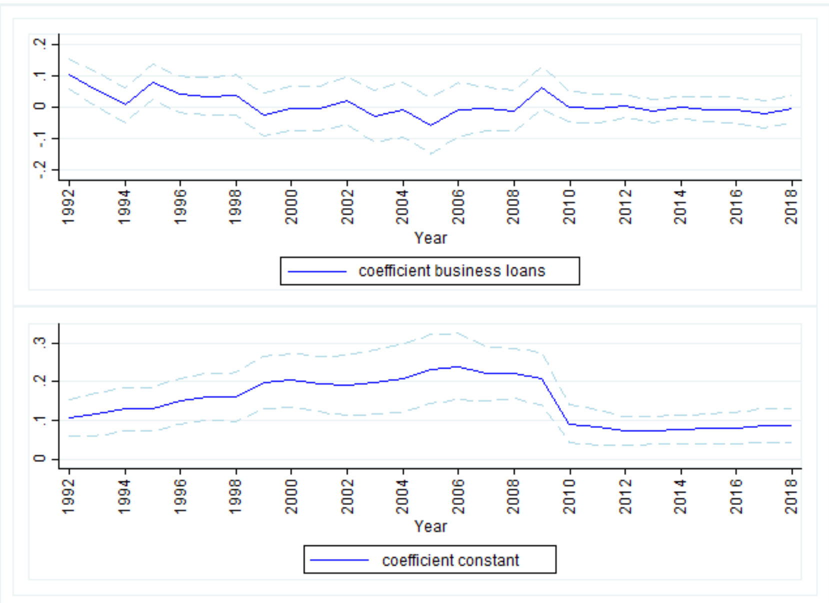


Figure 4.9: Regression coefficients: big banks

For the small (and the medium sized) banks the general pictureis quite similar to the aggregate estimates. The average coefficient on business loans is small but positive and higher before 2010 (0.17 on average) than after 2010 (0.06 on average). The constant trends upwards from 0.06 in 1992to 0.15 in 2009 and then drops to under 0.05 in 2010, where the confidence band includes zero for most of the years after 2010. For the big banks however, the picture is different. Except in 1992 the coefficient on business loans is not significantly different from zero anymore (the coefficient of the constant is similar to the other estimates). This is also true for the very big banks. Thus for big banks variations in business loan ratios are unconnected to variations in volatile liabilities.

So we find a small but positive correlation between opaque assets (business loans) and our measure of (uninsured) demandable liabilities for small and medium sized banks up to the 75th percentile but no correlation for the 25% biggest banks. From the perspective of our model it seems that the disciplining role of demandable liabilities is more important for smaller banks although the magnitude is small. The missing link for larger banks could be interpreted in the sense that these banks might enjoy insurance beyond deposit insurance in the form of implicit too-big-to-fail guarantees. Investors will take this into account and this reduces - similar to deposit insurance - their incentives to monitor and “discipline” the banks.

Bibliography

Anat Admati and Martin Hellwig. *The Bankers' New Clothes: What's Wrong with Banking and What to Do about It-Updated Edition*. Princeton University Press, 2014.

Charles W Calomiris and Matthew Jaremski. Stealing deposits: Deposit insurance, risk-taking, and the removal of market discipline in early 20th-century banks. *Journal of Finance*, 74(2): 711–754, 2019.

Charles W Calomiris and Charles M Kahn. The role of demandable debt in structuring optimal banking arrangements. *American Economic Review*, pages 497–513, 1991.

Tri Vi Dang, Gary Gorton, Bengt Holmström, and Guillermo Ordonez. Banks as secret keepers. *American Economic Review*, 107(4):1005–29, 2017.

Asli Demirgüç-Kunt and Harry Huizinga. Market discipline and deposit insurance. *Journal of Monetary Economics*, 51(2):375–399, 2004.

Douglas W Diamond. Financial intermediation and delegated monitoring. *Review of Economic Studies*, 51(3):393–414, 1984.

Piet Eichholtz, Nagihan Mimirogluz, Steven Ongena, and Erkan Yonder. Distance eects in cmbs loan pricing. *Unpublished*, 2019.

João Granja, Christian Leuz, and Raghuram Rajan. Going the extra mile: Distant lending and credit cycles. Technical report, National Bureau of Economic Research, 2018.

Bengt Holmstrom and Jean Tirole. Financial intermediation, loanable funds, and the real sector. *The Quarterly Journal of Economics*, 112(3):663–691, 1997.

Òscar Jordà, Moritz Schularick, and Alan M. Taylor. The great mortgaging: housing finance, crises and business cycles. *Economic Policy*, 31(85):107–152, 01 2016. ISSN 0266-4658.

Appendix D

D.1 Thresholds for $S_{4,L}$

Proof. $S_{4,L}$ is feasible for any α if it is feasible at $\alpha = 0$. This implies threshold $\bar{\pi}(S_{4,L})$ solves

$$\bar{z}(S_{4,L})|_{\alpha=0} = \bar{\pi}(S_{4,L})A(qy_l + p_g(1-q)y_h) + (1-\bar{\pi}(S_{4,L}))r = 1$$

$$\leftrightarrow$$

$$\bar{\pi}(S_{4,L}) = \frac{1-r}{A(qy_l + p_g(1-q)y_h) - r} \tag{D.1}$$

$S_{4,L}$ is never feasible if it is infeasible at $\alpha = 1$. This implies threshold $\underline{\pi}(S_{4,L})$ solves

$$\bar{z}(S_{4,L})|_{\alpha=1} = \underline{\pi}(S_{4,L})A(qy_l + p_g(1-q)y_h) + (1-\underline{\pi}(S_{4,L}))r = 1$$

$$\leftrightarrow$$

$$\underline{\pi}(S_{4,L}) = \frac{1-r}{A(qy_h + p_g(1-q)y_l) - r} \tag{D.2}$$

Since $q > 0.5$ from assumption (4.14) the denominator of (D.2) must be higher than the denominator of (D.1). This implies $\bar{\pi}(S_{4,L}) > \underline{\pi}(S_{4,L})$. The conditions on p_b and r which accompany (D.1) and (D.2) would be obtained by using $\pi = \frac{p_b+q-1}{p_b+p_g-1}$ from (4.4) in (D.1) and (D.2) and then solving for p_b which needs to be below 1.

The threshold $\pi^*(S_{4,L})$ solves $\alpha(S_0) = \alpha(S_{4,L})$ from (4.16) and (4.24):

$$\frac{1-Ay_l}{Aq\Delta y} = \frac{1 - r - \pi^*(S_{4,L})\big(qAy_l + (1-q)Ap_g y_h - r\big)}{\pi^*(S_{4,L})\Delta y A(q - p_g(1-q))}$$

$$\leftrightarrow$$

$$\pi^*(S_{4,L}) = \frac{q(1-r)}{(1-q)A\big(p_g q y_h + y_l(p_g - q)\big) + (q - (1-q)p_g) - qr} \tag{D.3}$$

Comparing (D.1) and (D.3) yields $\pi^*(S_{4,L}) < \bar{\pi}(S_{4,L})$ by assumptions (4.6) and (4.14). Comparing (D.1) and (D.3) yields $\pi^*(S_{4,L}) > \underline{\pi}(S_{4,L})$ by assumptions (4.7) and (4.14).

■

D.2 Proof of Lemma 1

Proof. We need to show:

$$\bar{\pi}(S_{4,L}) = \frac{1-r}{A(qy_l + p_g(1-q)y_h) - r} > \frac{1-r}{A((1-q)y_l + p_g q y_h) - r} \qquad = \pi(S_{3,L}) \tag{D.4}$$

$$\pi(S_{3,L}) = \frac{1-r}{A((1-q)y_l + p_g q y_h) - r} > \frac{1-r}{A(qy_h + p_g(1-q)y_l) - r} \qquad = \underline{\pi}(S_{4,L}) \tag{D.5}$$

$$\pi(S_{3,L}) = \frac{1-r}{A((1-q)y_l + p_g q y_h) - r} > \frac{1-r}{p_g A y_h - r} \qquad = \pi(S_{3,4,L}) \tag{D.6}$$

We only look at the denominators. The first inequality holds if $qy_l + p_g(1-q)y_h < (1-q)y_l + p_g q y_h$. This holds because $q > 0.5$, $Ay_l < 1$ and $p_g A y_h > 1$. The second inequality holds if $(1-q)y_l + p_g q y_h < q y_h + p_g(1-q)y_l$ which must also hold since $q > 0.5$ and $y_h > y_l$. The last equality holds if $(1-q)y_l + p_g q y_h < p_g y_h$ which also holds from $Ay_l < 1$ and $p_g A y_h > 1$.

Finally we need to show that $\pi^*(S_{4,L}) > \underline{\pi}(S_{3,L})$ if $p_g > p_g^*$, i.e.

$$\frac{q(1-r)}{(1-q)A\big(p_g q y_h + y_l(p_g - q)\big) + (q - (1-q)p_g) - qr} > \frac{1-r}{A((1-q)y_l + p_g q y_h) - r} \tag{D.7}$$

Multiplying this out and collecting terms for p_g yields

$$p_g\big((1-q)(1-Ay_l) + qAy_h(2q-1)\big) > q(1-2(1-q)Ay_l) \tag{D.8}$$

where the bracket after p_g must be positive from $Ay_l < 1$ and $q > 0.5$. Also the bracket on the right-hand side must be positive from $q > 0.5$. (D.8) then yields

$$p_g > \frac{q(1-2(1-q)Ay_l)}{(1-q)(1-Ay_l) + qAy_h(2q-1)} = p_g^* \tag{D.9}$$

■

D.3 Data sources and description

We obtain bank level data from the US Federal Deposit Insurance Corporation (FDIC). FDIC provides comprehensive quarterly bank level data collected through the *call reports*. Since we are interested in mainly stock variables, we work with the fourth quarter data for each of these years from 1992 through 2018. As the data is provided at the bank, we aggregate the data at the bank holding company (BHC) level for each year using the FRB ID Number for the Band Holding Companies ('rssdhcr'). The idea behind aggregation is that the decision on assets and liabilities structure are made at the BHC level rather than at the bank level. In the following we will refer to BHCs also as *banks*. We cleaned the data dropping bank-year pairs if the share of loan ratios are above 90% or below 0.1%. The idea is to drop banks that specialize to the extreme as most likely these would be banks created for special purposes by local of federal governments. After cleaning the data, we are left with a total of about 186,600 bank-year observations.

Volatile liabilities, according to the FDIC, include:

1. Time deposits that are *uninsured* and foreign office deposits

2. Federal funds purchased and repo borrowings

3. Demand notes issued to the US Treasury and other borrowed money with remaining maturity of 1 year or less, including Federal Home Loan Bank (FHLB) advances

4. Trading liabilities less trading liabilities revaluation losses on interest rate, foreign exchange rate, and other commodity and equity contracts.

The definition of *volatile liabilities* changed with effect from March 2010 as deposit insurance was expanded to cover deposits upto USD $250,000$ from USD $100,000$. Thus, uninsured deposits were redefined to all time deposits above USD $250,000$.

D.4 Estimates for other size classes

Figure D.1: Regression coefficients: medium sized banks

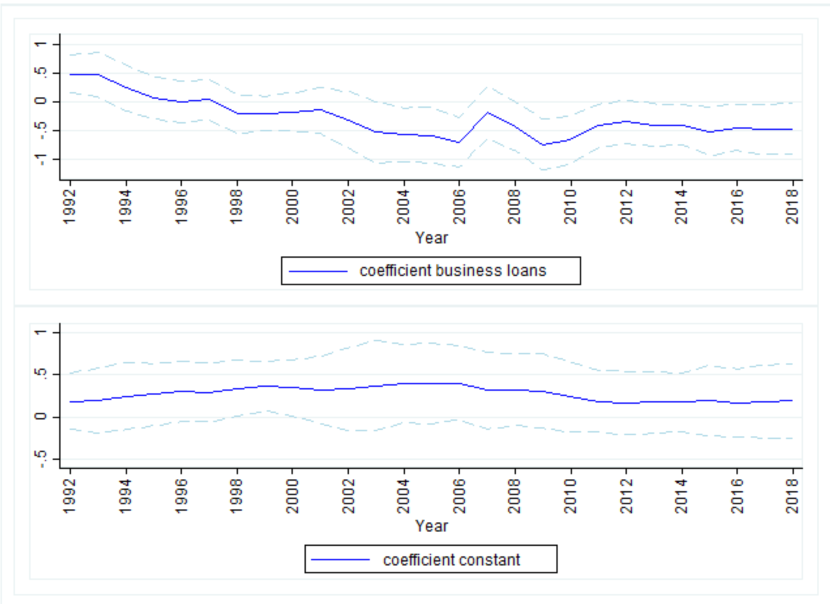


Figure D.2: Regression coefficients: very big banks