Should Banks Create Money?

Christian Wipf*

February 16, 2020

Abstract

The paper compares the welfare properties of two competing organizations of the monetary system: The current fractional reserve banking system where banks create inside money versus a narrow banking system where inside money creation is prohibited and the money supply consists of outside money issued by the central bank. Using a New Monetarist model, the analysis shows 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. Narrow banking provides no such compensation since inside money is fully backed by non-interest bearing outside money but it offers better insurance against liquidity risk. Welfare is higher under fractional reserve banking if banks can create a sufficiently high amount of inside money. (JEL: E42, E51, G21)

*University of Bern, Schanzeneckstrasse 1, P.O. Box 8573, CH-3001 Bern, Switzerland, christian.wipf@vwi.unibe.ch. I thank Cyril Monnet, Aleksander Berentsen, Dirk Niepelt, Lukas Altermatt, Mohammed Ait Lahcen and Lukas Voellmy for discussions and comments.

1 Introduction

Contemporary monetary systems are characterized by a mixture of public and private provision of 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 commercial banks issue inside money in excess of the outside money they hold against redemptions, i.e. they create inside money and we have a fractional reserve banking system.[1] The desirability of this financial structure has been debated throughout history and the debate revived during the recent financial crisis.[2] Opponents typically argue that fractional reserve banking brings instability without social benefits. Inside money creation should thus be prohibited by a separation of the monetary function of banks from their other functions: if banks issue inside money they must back it fully with very safe and liquid assets while lending and investing must 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 narrow banking plan was advocated by Friedman [1960] and related proposals were discussed after the recent financial crisis.[3]

The paper analyzes the persuasiveness of such proposals, focusing on the potential benefits of fractional reserve banking that arise from inside money creation.[4] A useful starting point is the quantity theory which predicts that the quantity of money and its composition between inside and outside money is irrelevant for the real allocation and welfare. Why should the quantity theory not apply when banks create inside money? The paper shows that when holding money is costly and outside money pays no interest the quantity of inside money w.r.t. outside money matters positively for welfare. The reason are 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.

This argument is developed in a standard "New Monetarist" model in the style of Lagos and Wright [2005] as presented in Rocheteau and Nosal [2017]. There

[1]The main reasons are that holders of demand deposits don't redeem all their demand deposits at once and that banks net the the in- and outflows of demand deposits from transactions vis-a-vis other banks over a clearing system. I use the terms "inside money creation" and "fractional reserve banking system" interchangeably for a situation where commercial banks issue more inside money than they hold outside money, i.e. they influence the money supply.

[2]In Switzerland there was even a vote on a ban of inside money creation in June 2018.

[3]See Pennacchi [2012] for an overview.

[4]The instability of fractional reserve banking systems has been extensively studied in the (microeconomic) banking literature and is quite well understood. The potential benefits of inside money creation have received much less attention. The macroeconomic literature where a fractional reserve banking system (with inside money being a claim on outside money) is compared to a narrow banking system is thin despite many papers on inside or private money. Papers most closely related to this are Sargent and Wallace [1982], Cavalcanti et al. [1999] and Chari and Phelan [2014]. The microeconomic literature is reviewed by Pennacchi [2012].

are consumers (buyers) and producers (sellers) and the buyers acquire money to buy goods from the sellers. There are two kinds of money: outside money issued by the central bank and inside money issued by perfectly competitive commercial banks. While outside money pays no interest, inside money is an interest-bearing claim on outside money and banks issue it against deposits in outside money (to partially back the inside money) and as loans. A crucial assumption is that the long run inflation rate lies above the Friedman rule. Hence, acquiring money is costly and buyers acquire too little relative to the social optimum. Inflation acts like a tax on real activity and this tax constitutes the fundamental source of inefficiency in the model. Since inside money is a claim on outside money the inflation tax affects both types of money.

In a first step (section 3) the only role of banks is to provide of liquidity, i.e. to issue inside money. I show that inside money creation improves the allocation. Fractional reserve banking is essential because inside money bears interest. If more inside money is issued (w.r.t. outside money) this means more lending, higher interest payments on bank assets (loans) and under perfect competition also on bank liabilities (inside money). Higher interest payments on inside money induce the sellers to produce more for the same price (or to produce the same for a lower price) and this leads buyers to acquire more inside money ex ante. Thus the 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 is a weighted average of the bank assets with zero nominal return (outside money) and a return equal to the inflation tax (loans) which is always below the inflation tax. The mechanism is identical to an economy where the central bank would pay interest on outside money.[5] The result is reminiscent of Chari and Phelan [2014] where more inside money is also beneficial because of higher interest payments. However, contrary to their model, the result here has nothing to do with insurance against preference shocks. Also, there is no pecuniary externality of inside money over the price level as in their model although agents are liquidity constrained as well.[6]

In a second step (section 4) I introduce another role for banks besides creating liquidity, namely the provision of liquidity insurance. This is accomplished by randomizing the division between buyers and sellers, i.e. by introducing a preference shock. The uncertainty whether agents will be buyers or sellers aggravates the basic inefficiency due to the inflation tax since now all agents acquire money. The risk of ending up as a seller with costly idle liquidity makes acquiring money even less attractive ex ante. In this situation the reallocation of liquidity through banks (by intermediating liquidity from sellers to buyers) after the shock can help because sellers can deposit their idle liquidity and buyers can borrow additional funds. Banks can provide this service both in a narrow banking and a fractional reserve system.

[5] On this see [Rocheteau and Nosal, 2017], chapter 6.2.

[6] Cavalcanti et al. [1999] also find that inside money creation improves welfare if outside money is sufficiently scarce. However, there are no prices and no inflation in their model so the welfare results seem to be driven mainly by the availability of inside or outside money.

Now both banking systems are beneficial and for some parameter values welfare in the fractional reserve banking system is higher while for others welfare in the narrow banking system is higher. In the narrow banking economy banks intermediate outside money from sellers to buyers and the rate on deposits equals the inflation tax. Agents who turn out to be sellers and deposit their money are thus fully compensated for incurring the inflation tax. As a consequence the narrow banking economy achieves full insurance against the preference shock. However, the basic inefficiency due to the inflation tax is not addressed since outside money pays no interest. In the fractional reserve economy, the interest rate for cash deposits is below the inflation tax. Thus the insurance against liquidity risk is imperfect. However, the basic inefficiency due to the inflation tax is partially addressed by interest on inside money. Narrow banks reallocate liquidity more efficiently and provide better insurance against the preference shock while fractional reserve banks create welfare increasing (interest bearing) liquidity. If fractional reserve banks create a sufficiently high amount of inside money (i.e. the interest on inside money is sufficiently high) the latter advantage outweighs the former disadvantage.

It is interesting to relate these results to the way both systems generate and distribute interest payments. In the narrow banking economy the generated interest is fully targeted to insure agents against the preference shock. An interest rate equal to the inflation tax is paid on the amount of outside money deposited by sellers in banks. In the fractional reserve economy on the other hand, interest is equally distributed over the full stock of inside money, but at a rate below the inflation tax. Comparing welfare results in a narrow and a fractional reserve banking economy where the interest payments per unit of money are identical, it turns out that the allocation in the fractional reserve economy is better. It is more efficient to pay a lower interest on the whole stock of money than a higher, fully-compensating interest on the deposits by sellers only. This is in the spirit of the findings by Sargent and Wallace [1982].

The rest of the paper is organized as follows: Section 2 shows the basic environment. Then the the basic model with inside money creation (section 3) and the model with inside money creation and liquidity insurance (section 4) are presented.

2 environment

basic structure: The environment follows a standard model in the style of Lagos and Wright [2005] as presented in Rocheteau and Nosal [2017] chapter 3.3. Time is discrete and continues forever. Every period is divided into two sequential competitive markets called *first* and *second* market.[7] There is a perishable consumption good produced and consumed in both markets denoted q in the first market and x in the second.

[7] Since both markets are competitive I follow Berentsen et al. [2007] and call these markets first and second market instead of the more common notations *DM* and *CM*.

agents: There is a unit mass of infinitely lived agents. Agents are of two types: there is a mass $s < 1$ of sellers and a mass $1 - s$ of buyers.[8] They discount future periods with β and they cannot commit. In the first market 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 they feature the quasi-linear utility function $U(x) - h$. $U(x)$ is strictly concave in x and also satisfies the Inada-conditions.

outside money, monetary policy and prices: There is a stock M of outside fiat money called "cash" issued by the central bank. The cash supply evolves at deterministic rate $\gamma > 0$, i.e. $M_t = \gamma M_{t-1}$. 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 or taxes τ to agents in the second market. Since agents have unit mass the transfer/tax per agent solves $\tau = M - M_{-1} = (\gamma - 1)M_{-1}$.

Let p be the price of consumption good q in terms of money in the first market and let P be the price of consumption good x in terms of outside money in the second market. The value of one unit of outside money in the second market is then the inverse of the price level $\phi = \frac{1}{P}$. Also denote (gross) inflation π as the ratio of the prices between two consecutive second markets, i.e. $\pi = \frac{P}{P_{-1}}$.

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 to agents. I assume bank debt (inside money) has the following properties: a) banks issue it in the second market and it is matures in the next second market[9] b) it is a claim on cash and pays interest (rate i_d) and c) it has the same liquidity properties as outside money and thus can circulate as inside money in the first market. I assume banks have to back the inside money they issue at least with a fraction α of outside money.

The idea behind this constraint is that transactions with inside money in the first market generate in- and outflows of inside money between banks and to settle and clear these flows banks need some outside money (think of reserves). The constraint says how much outside money an average bank needs at least to conduct these transactions. If for example a bank issued n units of inside money she needs at least αn units of outside money. The interpretation of α is less that of a policy parameter (reserve requirement). Rather, it captures how sophisticated and efficient the settlement/clearing system of an average bank or the banking system of a country works and is related to the technological development of the financial infrastructure.

[8] This is to simplify comparison with section 5 where agents are uncertain about their types, i.e. there is a preference shock which divides agents into buyers and sellers.

[9] With linear disutility of production in the second market there is no gain from spreading the redemption of inside money over multiple periods. Thus assuming this kind of contracts is not constraining in this environment.

the role for money and banks: The role for money in this environment is motivated by limited commitment. Since agents cannot commit, 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.

The only role of banks in this environment is that they can issue inside money. There is no liquidity-insurance or investment problem where they could help (the first will be changed in section 4). The question is whether such banks can improve the allocation compared to an economy with only outside money i.e. whether they are essential. The equilibrium allocation of an economy with only outside money is presented in appendix A.1.

equilibrium: Throughout the paper I focus on stationary and symmetric equilibria since the focus is on the long-run. In a stationary environment the value of real balances is constant over time implying

$$\frac{M}{P} = \frac{M_{+1}}{P_{+1}}.$$

Since the stock of cash grows at γ also the price level must grow at γ. So the growth of the stock of cash must equal the inflation rate or $\pi = \gamma$. By setting the long-run growth rate of the stock of cash the central bank can thus also determine long-run inflation. Thus γ can also be interpreted as the long run inflation target of the central bank.

Throughout the paper I assume that this long-run inflation target is higher than the discount factor, i.e. $\gamma > \beta$ and the central bank thus does not follow the Friedman rule. This assumption is central as it introduces a basic inefficiency into the environment. If there is inflation above the discount factor holding money is costly (money has a negative net return). As a consequence agents will hold too little money for first best consumption in the first market. Inflation acts like a tax on consumption/production in the first market (i.e. there is an inflation tax). Since inside money is a promise on outside money the inflation tax also applies to inside money.

first best: In a stationary environment a social planer would maximize expected lifetime utility of all buyers and a sellers:

$$\max_{q,x} \quad \frac{1}{1-\beta}[(1-s)u(q_b) - sc(q_s) + (U(x) - x)]$$
$$s.t. \quad (1-s)q_b = sq_s$$

The first best stationary allocation thus satisfies:

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

3 A basic model of inside money creation

3.1 banks and inside money

As described in the introduction banks issue inside money in the second market. What do they acquire with it? Since they must hold at least a fraction α of the inside money they issue in outside money (reserve constraint)[10] they want to acquire outside money in the form of cash deposits. The only other asset they can acquire are loans from agents. Similar to inside money I assume bank loans are made in the second market and paid back in the next second market with interest rate i and denominated in outside money. A representative bank thus maximizes the value of her assets (cash and loans) minus the value of liabilities (issued inside money) in the second market s.t. having enough cash to satisfy the reserve constraint. If d are the cash deposits and l are the loans per client (in equilibrium this will only be buyers) the problem of a representative bank is:

$$\max_{l,d} = d + l_b(1+i) - (l+d)(1+i_d) \tag{2}$$
$$s.t. \quad \alpha(l+d) \leq d \tag{3}$$

If there is an interest rate spread ($i > i_d$) and if the interest on inside money is positive ($i_d > 0$) the bank would like to make a loan as big as possible and to have as little cash deposits as possible. Conjecturing that this will be true in equilibrium the reserve constraint will bind and the bank will not hold excess-reserves. Denoting total inside money issued per client as $n = l + d$ this means the share of cash in the bank balance sheet d/n is just α and the share of loans l/n is $1 - \alpha$. We get the following relationship between loans and cash deposits

$$l = \frac{1-\alpha}{\alpha} d \tag{4}$$

We see that the loan size increases exponentially as α decreases for a given d. If α is 0.5 the bank can make loans of d but if $\alpha = 0.25$ the bank can make loans of $3d$. Where does this exponential growth come from? Suppose the bank wants to increase the loan size by 1. How much more cash does the bank need

[10] In the appendix A.2 I present a model where this constraint arises from redemptions from inside to outside money before the first market.

to back this loan expansion? Suppose that $\alpha = 0.5$ i.e. for every issued unit of inside money the bank needs 0.5 units of cash. So if loans increase by 1 the bank needs 0.5 additional units of cash in a first step. But if this cash is deposited the bank issues additional 0.5 units of inside money. To back *this* inside money the bank needs again a fraction α of it in cash so $0.5 * 0.5 = 0.25$ units. But if this cash is deposited the bank again issues additional inside money and so on and in the end the initial increase in loans from 1 triggers a cash need of $0.5 + 0.25 + 0.125 + ... = 1$ units of cash. If the bank issues one unit of inside money as a loan this triggers an infinite sum of subsequent cash acquisitions who converge to $\frac{\alpha}{1-\alpha}$. Clearly this sum is exponentially lowered if α decreases. For example if $\alpha = 0.25$ the bank needs only $1/3$ units of cash if she wants to increase the loan size by one. This exponential decrease in the necessary cash holdings when α decreases gives rise to the exponential increase in the loan size.

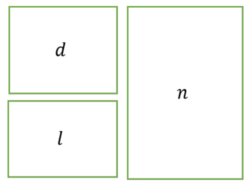
Using the binding reserve constraint we can rewrite the objective function as:

$$\max_{n} \quad n(\alpha + (1-\alpha)(1+i) - (1+i_d))$$

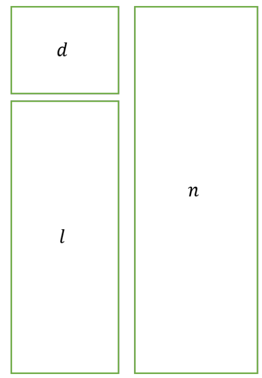
In equilibrium free entry implies banks will enter until they earn zero profits and thus the following relationship between interest on inside money and the loan rate must hold in equilibrium:

$$1 + i_d = \alpha + (1-\alpha)(1+i) \tag{5}$$

So the interest rate on inside money is a weighted average of the return on cash (1 in nominal terms) and the return on loans $(1+i)$. Consequently the interest on inside money must be below the loan rate so there is a spread if $\alpha > 0$ (the first conjecture was correct). We also see that the interest on inside money negatively depends on α. Remember that α denotes the share of cash in the bank balance sheet d/n. So if α increases the share assets with no return (cash) in a bank's balance sheet increases and the share of assets with return (loans) decreases. Consequently the bank can pay less interest on its liabilities. Both effects are shown graphically in the figure below.



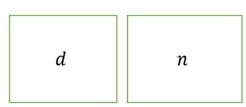
$i \ggg i_d$, big α



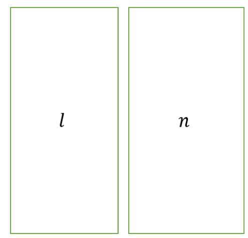
$i > i_d$, small α

In the left case α is big. The bank needs to back the issued inside money with a lot of cash and as a consequence the amount of loans she can make is small. This bank has a lot of assets with no return (cash) and relatively little assets with returns (loans). So in this case we have a lower interest on inside money (taking the interest on loans as given) and a bigger spread between loan rate and the rate on inside money. In the right case α is low. the bank has a lot of assets with returns (loans) compared to assets with no returns (cash). Therefore interest rates on loans and inside money will be much closer together (lower spread) and the interest on inside money is higher. If α is so small that the share of cash becomes negligible the bank would pay practically the same interest rates on loans and banknotes.

For illustration and completeness I also consider the corner cases when $\alpha = 1$ and $\alpha = 0$. If $\alpha = 1$ this means that the bank must store all cash deposited. This bank cannot make any loans and the balance sheet just consists of cash and inside money issued against cash deposits. Since cash does not pay any (nominal) return the bank cannot pay any interest on the inside money issued and under zero profits $i_d = 0$. This would be a narrow banking economy.



On the other hand if $\alpha = 0$ the bank would need no cash to issue inside money. If the interest rate on banknotes is positive banks don't hold cash anymore and they would only issue loans. Their assets would then consist of only loans. In this case zero profits implies that the loan rate and the rate on banknotes are equal, i.e. $i = i_d$.



3.2 second market

We now turn to the question how much money agents want to hold. I make two simplifying conjectures which will be verified in equilibrium: First I conjecture that if holding money is costly only buyers will acquire it. I thus focus on a representative buyer problem in the second market. Second, I focus on the case where buyers only acquire inside money and no outside money. In the appendix A.3 I show that buyers will do this if the interest on inside money (i_d) is positive (so this is in line with the conjecture already used in the banking problem). Note that if buyers only hold inside money this also implies that only banks hold outside money in this economy.

As laid out above a buyer has two possibilities to acquire inside money in the second market. She can deposit cash in the bank (denoted d) which the bank needs to fulfill the reserve requirement or she can borrow inside money (denoted l). If agents borrow they get inside money now and pay back in the next second market with interest i. So either agents work today to acquire inside money or they work tomorrow. The amount of inside money an agent brings into the next period is the sum of cash deposits and loans i.e. $n_{+1} = d + l$. The problem of a buyer arriving in the second market with n units of inside money and l_{-1} debt from last period is thus:

$$\begin{aligned} W(n, l_{-1}) &= \max_{x,h,d,l} U(x) - h + \beta V(\overbrace{d + l}^{=n_{+1}}, l) \qquad (6) \\ s.t. \quad x + \phi d &= h + \phi\tau + n\phi(1 + i_d) - l_{-1}\phi(1 + i) \end{aligned}$$

The first-order condition for optimal consumption in the second market is $U'(x^*) = 1$ (this also applies to sellers)[11] and optimal cash deposits and optimal borrowing solve:[12]

$$\phi = \beta V_{n+1}(n_{+1}, l) \qquad (7)$$

$$V_l(n_{+1}, l) = V_{n+1}(n_{+1}, l) - \phi_{+1}(1 + i_{+1}) = 0 \qquad (8)$$

[11] For this conclusion to hold we need to apply a scaling condition to $U(..)$ s.t. $x^* = U'^{-1}(1)$ is big enough that the amount sellers work h_s is always positive, see [Berentsen et al., 2007] proof of proposition 1. I will assume such a condition is always satisfied.

[12] Since any meaningful equilibrium must have positive cash holdings and positive loans by banks I focus on interior solutions. The partial derivatives of the value functions are denoted with subscripts, e.g. $\frac{\partial V(n_{+1},l)}{\partial n_{+1}} = V_{n+1}(n_{+1}, l)$.

This means buyers must be indifferent between both forms of acquiring inside money. As a consequence in any equilibrium the loan rate must equal the inflation tax or:

$$\frac{\phi}{\phi_{+1}}\frac{1}{\beta} = 1 + i_{+1} \tag{9}$$

Finally, the value of bringing an additional unit of inside money in the second market and the marginal value of borrowing last period are:

$$W_n(n, l_{-1}) = \phi(1 + i_d) \tag{10}$$
$$W_{l_{-1}}(n, l_{-1}) = -\phi(1 + i) \tag{11}$$

And the market clearing conditions for the output good, cash deposits and borrowing are:[13]

$$(1 - s)h_b + sh_s = (1 - s)x_b + sx_s = x^* \tag{12}$$
$$(1 - s)d = M \tag{13}$$
$$(1 - s)l = (1 - s)\frac{1 - \alpha}{\alpha}d = \frac{1 - \alpha}{\alpha}M \tag{14}$$

3.3 first market

With inside money used in the first market a representative buyer with n units of inside money solves the following problem:

$$V_b(n, l_{-1}) = \max_{q_b} \quad u(q_b) + W(n - pq_b, l_{-1})$$
$$s.t. \quad pq_b \leq n$$

Using that the marginal value of inside money is $\phi(1+i_d)$ from (10) and defining q^* as the quantity consumed if the liquidity-constraint is slack (i.e. q^* solves $u'(q^*) = p\phi(1 + i_d)$) the solution for optimal consumption and the marginal value of inside money is:

$$\{q_b, V_{b,n}(n, l_{-1})\} = \begin{cases} \{q^*, \phi(1 + i_d)\} & \text{if} \quad n \geq pq^* \\ \\ \{\frac{n}{p}, \frac{u'(n/p)}{p}\} & \text{if} \quad n < pq^* \end{cases} \tag{15}$$

A representative seller with no inside money holdings (and also no borrowing) solves:

[13] The first expression uses the fact that both buyers and sellers will consume the same amount x^* in the second market. The second expression uses the conjecture that only buyers deposit outside money and in the end all outside money will be held by banks. And the third expression uses the binding reserve constraint of the bank (56).

$$V_s(0,0) = \max_{q_s} \quad -c(q_s) + W(pq, 0)$$

with the solution:

$$c'(q_s) = p\phi(1 + i_d) \tag{16}$$

Since inside money pays interest sellers incorporate this in their working decision.

Finally market clearing in the first market is given by:

$$sq_s = (1 - s)q_b \tag{17}$$

3.4 equilibrium

To solve for the stationary equilibrium allocation I conjecture that a buyer is liquidity constrained if $\gamma > \beta$ i.e. that we are in the lower case of (15). Optimal inside money holdings of buyers then solve:

$$\phi = \beta V_{n+1}(n_{+1}, l) = \beta \frac{u'(q_{+1})}{p_{+1}} \tag{18}$$

If we combine this with the seller optimality condition (16) and apply $\gamma = \phi/\phi_{+1}$ (stationarity) we can get equilibrium consumption in the first market as a function of the inflation tax and the interest on inside money:

$$\frac{u'(q_b)}{c'(q_s)} = \frac{\gamma}{\beta} \frac{1}{1 + i_d} \tag{19}$$

Note that the expression is identical to the allocation in an economy without banks (A.1) except for the interest rate on inside money. And also note that first market consumption depends positively on this interest rate. The closer the interest on inside money to the inflation tax, the closer consumption in the first market to first best consumption q^*.[14]

Next I derive equilibrium interest rates. To get the equilibrium loan rate I apply $\gamma = \phi/\phi_{+1}$ to the indifference condition between both forms of inside money (9). The loan rate must equal the inflation tax and is thus independent of α. The

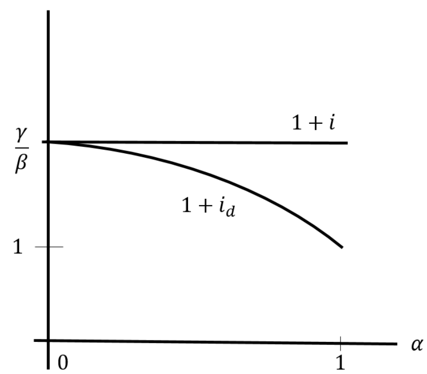
[14] Under the assumptions made on preferences, $\frac{u'(q)}{c'(q)}$ decreases in q so if the RHS of (19) is bigger than 1 it must be that $q_b < q_b^*$ (see A.1). This logic applies to all the comparative statics in the models used. Every change in parameters which brings the RHS closer to 1 is welfare improving because it gets closer to the first best allocation q^* where $\frac{u'(q^*)}{c'(q^*)} = 1$.

interest on inside money is then the weighted average of the loan rate and the return on cash from the zero profit condition of banks (5) and is thus decreasing in α. Thus equilibrium interest rates are

$$1 + i = \frac{\gamma}{\beta} \tag{20}$$

$$1 + i_d = (1 - \alpha)\frac{\gamma}{\beta} + \alpha \tag{21}$$

and we get the following picture for the evolution of the interest rates as a function of α:



We saw that α is the share of cash to total bank assets (and $1-\alpha$ is the share of loans). So if α increases this means the share of assets with no nominal return (cash) increases vis a vis the share of assets with a positive nominal return (loans). In equilibrium the return on loans is independent of α so consequently the bank can pay less interest on its liabilities if the asset-mix shifts towards assets with no return (cash). As described in the section on banks, an increase in α leads to a very strong increase in the amount of cash banks need to hold if they want to make additional loans. Thus lending reacts very strongly to changes in α.

To get equilibrium consumption in the first market I combine 19 with the equilibrium expression for the interest on inside money to get and market clearing 17:

$$\frac{u'(q_b)}{c'(\frac{1-s}{s}q_b)} = \frac{\gamma/\beta}{(1-\alpha)\gamma/\beta + \alpha} \tag{22}$$

The following proposition summarizes the most important results:

Proposition 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 and inside money in which:*

i) first market consumption solves (22)

ii) first market consumption is below first best q^* and the inefficiency increases in the inflation tax γ/β and the reserve constraint α.[15]

iii) the allocation is better than in an economy without banks (see A.1), i.e. banks 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 as in A.1.

v) as $\alpha \to 0$ outside money is not used anymore and the allocation approaches the first best allocation (1).

Several results of this proposition are worth highlighting. First, the proposition shows a non-neutrality result with respect to inside money (*ii*)). The more inside money relative to cash (the lower α) the higher welfare. Thus the quantity theory does not hold with respect to inside money in this economy. This implies, second, that banks with the only function of inside money creation are essential, i.e. they improve the allocation compared to an economy without banks and inside money creation (*iii*)). I will explain these results in more detail in the next section. The proposition also says that if banks don't have a liquidity creating function, i.e. if $\alpha = 1$ and they are narrow banks, the allocation is the same as without banks (*iv*)).We saw that if $\alpha = 1$ banks cannot offer any loans and they cannot pay interest on inside money either. This implies buyers cannot borrow and there is no benefit from depositing cash in banks compared to holding cash directly since both have the same liquidity properties. In fact inside and outside money are identical in such an economy and thus the allocation must be equivalent to an economy without banks where buyers directly acquire cash. Narrow banks offer no social benefits and are not essential in this environment. If the only role of banks is inside money creation and if narrow banking prohibits inside money creation it should be that a narrow banking system is equivalent to a system without banks.

Finally a few comments on the equilibrium if $\alpha = 0$ (*v*). In the section on banks we saw that for positive interest rates on inside money banks will not hold any cash and interest rates on loans and inside money will be equalized under zero profits, i.e. $i = i_d$. In an equilibrium where banks don't demand cash deposits, also the supply of cash deposits must be zero. Thus it must be that buyers want to borrow but they don't want to make cash deposits. The condition for this to happen is that the condition for borrowing in (7) is interior but the marginal

[15] It is a standard feature of models in the tradition of Lagos and Wright [2005] that second market consumption is always at first best. Thus for welfare comparisons between different models it is sufficient to focus on the first market. This holds also for the models in the next sections.

costs of depositing are higher than the marginal benefits i.e.:

$$\phi > \beta V_{n+1}(n_{+1}, l) = \beta\phi_{+1}(1 + i_{+1})$$

Thus in a stationary equilibrium buyers borrow but don't deposit if the loan rate is below the inflation tax:

$$\frac{\gamma}{\beta} > 1 + i. \tag{23}$$

Any (positive) interest rate satisfying 23 would be an equilibrium if $\alpha = 0$. In such an economy outside money has no function anymore. Because the interest on borrowing 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. $n \in (pq^*, \infty)$.[16] The economy achieves the first best allocation equivalent to an economy with direct credit.

3.5 the non-neutrality of inside money

Why is more inside money (a lower α) beneficial? Above I explained how a lower α leads to strong 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.[17]

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 (16), $c'(q) = p\phi(1 + i_d)$, and equilibrium demand is given by the buyer optimality condition $u'(q) = \gamma/\beta p\phi = (1+i)p\phi$. From the properties of the utility functions we can infer that supply (weakly) increases in the relative price $p\phi$ and in the interest inside money i_d and demand decreases in the relative price $p\phi$ and in the loan rate. From this the effects of the changes in interest rates through α on equilibrium production/consumption can be inferred. We saw above that a decrease in α *increases* interest on inside money thus supply shifts up. Sellers want to produce more for the same relative price because they get more interest in the next second market. The following figure depicts this shift from initial supply $S(\alpha_1)$ to the new supply curve $S(\alpha_2)$ for a decrease in α from an arbitrary

[16] The definition of inside money as a claim on outside money may seem strange in an economy where only inside money is used

[17] 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 expression is exactly identical to equation 22.

value α_1 to a lower value α_2. The figure shows that as a result of this change equilibrium consumption (and production) will clearly increase.

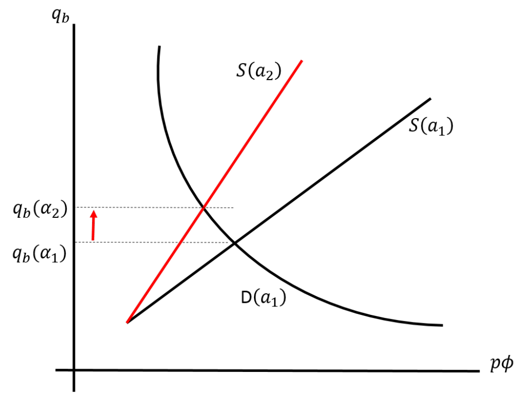


Figure 1: The equilibrium effects of an increase in interest of inside money

How can we see this non-neutrality from a quantity theory perspective? Look at the inside money $n = l + d$ a buyer brings to the first market. In equilibrium no cash is held outside banks so the amount of cash deposits in banks must equal the stock of cash in the economy or $d = M/(1-s)$ from (13). Therefore the inside money holdings of a buyer are $n = \frac{M}{\alpha(1-s)}$. Equilibrium consumption in the first market can then be written as:

$$q_b = \frac{n}{p} = \frac{M/(\alpha(1-s))}{p} \tag{24}$$

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 n doubles. If the quantity theory would hold, we would expect the price level to increase by the same factor and consumption would not change. However, this is not what happens in this economy. We saw that over the changes in interest rates a reduction in α leads to an increase in q_b. So equation 24 can only hold if the price level in the first market rises *less* than the amount of inside money available $(M/(\alpha(1-s))$.

4 Inside money creation and liquidity insurance

In the preceding section we saw that the inside money creation is beneficial because of the interest on inside money. In this section I want to ad another role for banks besides creating liquidity, the provision of liquidity insurance, and see how this affects the results from section 3.

4.1 changes in the environment

liquidity risk: Following Berentsen et al. [2007] (BCW in the following) a liquidity insurance role for banks can be introduced by two simple modifications to the environment: The first is to make the division into buyers and sellers random from the perspective of agents, i.e. to introduce a *preference shock*. Now agents ex-ante don't know whether they will be buyers or sellers in the first market. With probability s an agent is a seller in the first market and with the inverse probability $1 - s$ an agent is a buyer. If acquiring liquidity is costly (which is the case here since $\gamma > \beta$) then there is room for the reallocation of liquidity through banks after this shock. But to make this possible the interaction of banks and agents must be after the shock. This is done by the insertion of a *banking period* after the preference shock. Contracts between banks and agents (bank debt and loans) are now formed in this banking period and are redeemed in the second market (before: from second market to second market). This is the second modification. Thus the sequence of events in one period is now as follows:

$t \rightarrow$ preference shock $\rightarrow$ banking period $\rightarrow$ first market $\rightarrow$ second market $\rightarrow t{+}1$

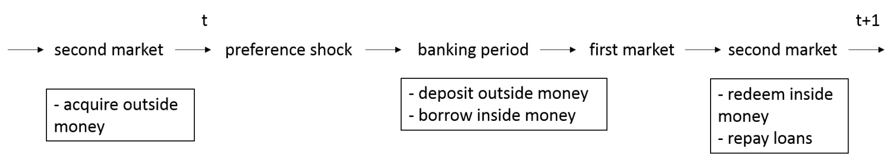
What are the implications of the introduction of this preference shock? In appendix A.4 I show that introducing a preference shock into an economy with only outside money worsens the allocation compared to the allocation without preference shock and only outside money in appendix A.1. If acquiring money is costly, the risk to be a seller with (costly) idle liquidity in the first market makes acquiring money even less attractive ex-ante. So the uncertainty from the preference shock aggravates the basic inefficiency from the inflation tax. The introduction of a preference shock has no formal implications for the first best allocation (1). The objective function of the social planer is now interpreted as *expected* lifetime utility of a representative agent under stationarity.

two banking regimes: Due to the introduction of liquidity risk banks can have two roles now: the creation of liquidity and the insurance against liquidity risk. How should we think of fractional reserve banking and narrow banking in this extended environment? In the introduction I defined a *narrow banking* regime as a banking system where inside money creation is prohibited but banks can perform other functions like intermediation if they don't finance them with inside money. In the model environment this is equivalent to an economy with only outside money where the money supply is fully controlled by the central bank.[18] Thus in the narrow banking regime there will be only outside money but banks can intermediate this from sellers to buyers. This economy is a basic version of BCW.[19] In contrast to this in a *fractional reserve* banking system banks operate as before: they issue inside money against cash deposits and loans subject to a reserve constraint. So in this economy banks have both roles: they provide liquidity and they provide liquidity insurance.

In the following I solve the model under each of these two regimes taking the regime as exogenous, i.e. I assume that banks and agents cannot decide on the regime. The basic question is again: is fractional reserve banking essential? Does it improve the allocation with regard to narrow banking?

4.2 Fractional reserve banking

Banks are now active from the banking period to the second market. This means agents can only acquire outside money in the second market. Then, in the banking period after the preference shock, agents can deposit this outside money and borrow. Again banks issue inside money against cash deposits d and loans l (at rate i) and inside money is redeemed with interest in the second market with interest i_d. The sequence of events is summarized in the following graph:



4.2.1 second market

A representative agent (which may have been a buyer or a seller in the first market before) may bring outside money (m) inside money n and some debt l into the second market. He chooses consumption x, work h and his new holdings of outside money m_{+1}. Note that now, since agents face the uncertainty of the preference shock, $V(m_{+1})$ denotes the *expected* value of entering the next period with m_{+1} units of outside money.

$$\begin{aligned} W(m,n,l) = & \max_{x,h,m_{+1}} U(x) - h + \beta V(m_{+1}) \qquad (25) \\ s.t. \quad & x + \phi m_{+1} = h + \phi(\tau + m) + n\phi(1+i_d) - l\phi(1+i) \end{aligned}$$

[18]The narrow banking allocation in the preceding section was the same as an economy with only outside money, see proposition 1 *iv*)

[19]In footnote 9 in BCW the authors also make the interpretation of their model as a narrow banking economy.

The first order condition for optimal consumption is again $U'(x^*) = 1$ and optimal (positive) outside money holdings solve:

$$\phi = \beta V'(m_{+1}) = (1-s)V_b'(m_{+1}) + sV_s'(m_{+1}). \tag{26}$$

Both first-order conditions imply that agents want to work such that they consume the same amount x^* in the second market and they choose the same amount of cash to bring into the next period - independent of m, d and l. These properties follow from the quasi-linear utility function introduced by Lagos and Wright [2005]. The expected marginal value of outside money is the marginal value of outside money as a buyer $V_b'(m_{+1})$ times the marginal value as a seller $V_s'(m_{+1})$ weighted with the respective probabilities. The envelope conditions to the problem are:

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

4.2.2 banking period and first market

Again I focus on the case where agents use only inside money in the first market referring to appendix A.3. Also, 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 (d_b, where $d_b \leq m$) 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.

$$\begin{aligned} V_b(m) = \max_{q_b, l_b, d_b} \quad & u(q_b) + W(m - d_b, l_b + d_b - pq_b, l_b) \\ s.t. \quad & pq_b \leq d_b + l_b \\ & d_b \leq m \end{aligned}$$

First it is clear that if inside money is used in the first market a buyer should deposit all his cash in the bank. No matter whether he is liquidity constrained in the first market (i.e. the first constraint binds) he is always better off depositing the cash if interest is positive. So we must have $d_b = m$. The solution to the problem is then.

$$
\{q,l\} = \begin{cases} \{q^*, \in (pq^*-m,\infty)\} & \text{if} \quad i = i_d, m < pq^* \\ \\ \{\frac{m+l}{p}, l>0 \quad \text{solves} \quad u'(\frac{m+l}{p}) = p\phi(1+i)\} & \text{if} \quad i > i_d, u'(\frac{m}{p})\frac{1}{p} > \phi(1+i) \\ \\ \{\frac{m}{p}, l=0\} & \text{if} \quad i > i_d, u'(\frac{m}{p})\frac{1}{p} \leq \phi(1+i) \end{cases} \tag{27}
$$

We have tree possible situations. In the first, the loan rate and the interest on inside money are the same, $i = i_d$. This means from the perspective of a buyer borrowing is costless and buyers can acquire as much inside money as they like. Consequently consumption is at first best q^* and borrowing is in the interval $(pq^* - m, \infty)$.[20] In the second situation borrowing is costly, i.e. the loan rate is above the interest on inside money or $i > i_d$ but the amount of outside money is sufficiently small (or borrowing is sufficiently cheap) that buyers still want to borrow.[21] In this situation a buyer will never have enough inside money for first-best consumption. In the third situation a buyer is also liquidity constrained but the amount of outside money is sufficiently high (or borrowing is so expensive) that he prefers bringing outside money and depositing it over borrowing. From the banking problem we already know that the loan rate must be above the interest rate on inside money (see above). So we are in situations two or three. Since any meaningful equilibrium should have positive loans I will also conjecture that the conditions for the second situation are satisfied in equilibrium and later verify that they hold.

In situation 2 the marginal value of cash for a buyer in the first market $V_b'(m)$ is equal to

$$
V_b'(m) = \frac{u'(\frac{m+l}{p})}{p} = \phi(1+i) \tag{28}
$$

seller problem

A representative seller solves the following problem in a RC-economy:

$$
\begin{aligned} V_s(m) = \max_{q_s, l_s, d_s} \quad & -c(q_s) + W(m - d_s, d_s + l_s + pq_s, l_s) \\ & s.t. \quad d_s \leq m \\ & \qquad l_s \geq 0 \end{aligned}
$$

[20]We still know from before that if acquiring money is costly, $\gamma > \beta$ a buyer will never bring enough outside money for first best consumption q^* so we must have $m < pq^*$.

[21]The second condition says that the marginal benefits of borrowing are higher than the marginal costs.

It can easily be shown that a seller will not borrow if $i > i_d$, so $l_s = 0$ and he will deposit all the outside money he brings if $i_d > 0$, so $d_s = m$. Also optimal production is again given by equalizing marginal disutility of work with marginal utility given by 16. The marginal value of outside money for a seller in the first market $V_s'(m)$ is equal to the marginal value of inside money in the next second market (since he deposits it).

$$V_s'(m) = \phi(1 + i_d) \tag{29}$$

Market clearing in the first market is still given by (17).

optimal cash holdings

We can iterate (28) and (29) one period forward and use them in (26) to get a more intuitive expression for the optimal holdings of outside money. The marginal costs of acquiring it must equal the weighted average of the two interest rates which denote the marginal utilities of outside money as a buyer and a seller.

$$\phi = \beta\phi_{+1}[(1 - s)(1 + i_{+1}) + s(1 + i_{d+1})]. \tag{30}$$

banks

The bank problem is essentially the same as in the economy without preference shock. As before I express the problem in per buyer terms. The only difference is that now also sellers deposit and thus the market clearing for cash deposits must also take into account the sellers and d now denotes the amount of cash deposits of a bank *per buyer* (and not of a buyer). Still we get a binding reserve constraint for $i > i_d > 0$ so (56) and (5) must hold here too. Again the interest on inside money is the average of the return on loans $(1 + i)$ and the return on cash weighted by the relative shares of these assets:

$$1 + i_d = \alpha + (1 - \alpha)(1 + i)$$

The market for loans and cash deposits clears as follows:

$$(1 - s)l = \frac{1 - \alpha}{\alpha}(1 - s)d = \frac{1 - \alpha}{\alpha}((1 - s)d_b + sd_s) \tag{31}$$

4.2.3 equilibrium

Using stationarity ($\gamma = \phi/\phi_{+1}$) on optimal outside money holdings of agents (30) we get that the weighted average of both interest rates must equal the inflation tax:

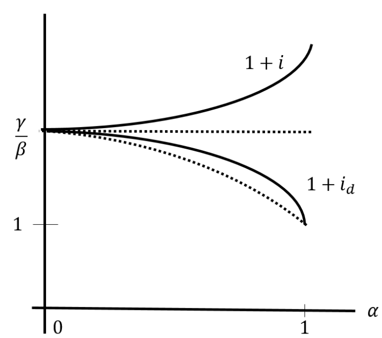
$$\frac{\gamma}{\beta} = (1-s)(1+i) + s(1+i_d) \tag{32}$$

Note that given a level of the inflation tax this optimality condition introduces a negative relationship between both interest rates. The higher the loan rate the lower the interest on inside money must be and the other way round, otherwise supply and demand for outside money would not be in balance. Now I combine this with the free entry condition of banks, (5), to get equilibrium interest rates:

$$1 + i_d = \frac{(1-\alpha)\gamma/\beta + \alpha(1-s)}{1-\alpha s} \tag{33}$$

$$1 + i = \frac{\gamma/\beta - \alpha s}{1-\alpha s} \tag{34}$$

Compared to the interest rates in the economy without preference shock we get two differences: The loan rate is now a positive function of α. The higher α the higher the equilibrium loan rate. This is a consequence of the negative relationship between loan rate and interest on inside money introduced by optimal outside money holdings (32). If α increases this decreases the interest on inside money payable given the loan rate over the zero profit condition of banks. But at this lower interest on inside money the demand for outside money would be 0 and thus the loan rate must rise. Second, comparing the interest on inside money in (33) with the analogue expression for the economy without preference shock (20) we see that for any given α the interest on inside money is higher in the economy with preference shock than without. The following figure shows these two differences (the dashed lines are the interest rates of the economy without preference shock).



Note that for $\alpha \in (0,1)$ interest rates always satisfy $i > i_d > 0$ and thus the conjectures for buyer, seller and bank behaviour were correct. To get equilibrium consumption in the first market I insert the expressions for equilibrium interest rates into optimal consumption in situation 2 in equation (27) and use optimal seller production (16).

$$\frac{u'(q_b)}{c'(\frac{1-s}{s}q_b)} = \frac{1+i}{1+i_d} = \frac{\gamma/\beta - \alpha s}{(1-\alpha)\gamma/\beta + \alpha(1-s)} \tag{35}$$

We get the following proposition for an economy with inside money creation and a preference shock:

Proposition 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 inside money creation in which:*

i) first market consumption solves (35)

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) the allocation is worse than in the inside money creation economy without preference shock in proposition 1.

iv) the allocation is better than without banks as in A.4 and approaches this allocation as $\alpha \to 1$.

v) as $\alpha \to 0$ outside money is not used anymore and the allocation approaches the first best allocation.

What are the new insights from this proposition compared to proposition 1? Let me first highlight the similarities. First, the non-neutrality result with respect to inside money is also present here (*ii*)). The more inside money (the lower α) the higher welfare. Second, we also get the result that as α goes to 0 the allocation is efficient (outside money is not used anymore). Also, fractional reserve banking is essential, i.e. the allocation is better than without banks (*iv*)). The new thing is the preference shock which matters negatively. The higher the fraction of sellers in the economy (or alternatively the higher the risk to be a seller) the lower welfare. I will give an interpretation for this below. Note that the preference shock is also the reason why the economy with preference shock is worse than without (*iii*)). As s goes to zero the allocation approaches proposition 1).

What can we say with regard to the comparison to a narrow banking regime? The proposition says that as α goes to 1 (all banks are narrow banks) the allocation is like without banks so banks have no socially useful role (*vi*)). However, we must be cautious before concluding that this is also the allocation of a narrow banking regime. A narrow banking regime prohibits only inside money creation but not the other functions of banks (as long as their liabilities don't circulate as inside money). Specifically, as shown in BCW, banks can still intermediate outside money from sellers to buyers in a narrow banking regime. In the next section I will characterize the allocation of such an economy. As it turns out, narrow banking is also essential with preference shock (the allocation is better than without banks) and the welfare comparison between fractional reserve and narrow banking depends on parameters.

4.3 Narrow banking

As I argued in the beginning of the section a narrow banking economy where inside money creation is prohibited is equivalent to an economy where only outside money is used.[22] The biggest changes are in the banking sector. Banks cannot issue inside money anymore so the only thing they can do is intermediate outside money from sellers to buyers. As BCW show in this economy sellers deposit cash in the banking sector (if interest rates on these deposits are positive) and buyers borrow cash. So we can write the cash profits of a representative bank in per buyer-terms as value of assets minus value of liabilities (d are again the cash deposits per buyer and l are the cash loans per buyer) subject to a feasibility constraint that the bank cannot make more loans than the cash deposited.

$$\max_{l_b,d} = l_b(1+i) - d(1+i_d)$$
$$s.t. \quad l_b \leq d$$

It is clear that the bank should lend out all cash deposited if $i > i_d$ (w.l.o.g. also if $i = i_d$). Free entry implies zero profits which equalizes the interest rates.

$$i = i_d \tag{36}$$

So in contrast to the economy with inside money creation there is no spread between the loan rate and the rate on cash deposits anymore. Finally market clearing for loans and deposits implies (already presuming that only sellers deposit):

$$(1-s)l_b = sd_s \tag{37}$$

4.3.1 equilibrium

As BCW show the condition for optimal cash holdings in the second market for agents is identical to the one in the fractional reserve banking system (32). So

[22] Since such an economy is a basic version of BCW I only highlight the most important differences to the fractional reserve banking economy and direct the interested reader to their paper for details.

combining this with the zero profit condition of banks (36) we get that both interest rates must equal the inflation tax or

$$1 + i = 1 + i_d = \frac{\gamma}{\beta} \tag{38}$$

Since I focus on the case where $\gamma > \beta$, interest rates are positive and thus buyers are liquidity constrained in the first market and sellers strictly prefer to deposit all their cash. As BCW (p. 180) show equilibrium consumption in the first market is given by:

$$\frac{u'(q_b)}{c'(\frac{1-s}{s}q_b)} = 1 + i = \frac{\gamma}{\beta}. \tag{39}$$

Proposition 3. *Suppose holding money is costly (i.e. $\gamma > \beta$) then there is a unique stationary equilibrium of a narrow banking economy in which:*

i) first market consumption solves (39)

ii) first market consumption is below first best consumption q^* and the inefficiency increases in the inflation tax γ/β.

iii) the allocation is identical to the allocation without preference shock and only outside money in appendix A.1.

iv) the allocation is better than the allocation with outside money and preference shock in appendix A.4 (no banks).

The proposition shows that narrow banking is essential (*iv*)). So what is the welfare gain from narrow banking? 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*)). The intuition is that since the deposit rate exactly equals the inflation tax ($1 + i_d = \frac{\gamma}{\beta}$) in the narrow banking economy 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 acquired before 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.[23]

4.4 Fractional reserves vs. narrow banking

In the preceding section it was established that in the economy with fractional reserves there is an equilibrium like the one without banks if $\alpha = 1$ and if $\alpha = 0$ the allocation is efficient. Also it was shown that welfare increases as α gets smaller. In the narrow banking economy it was established that the allocation is

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

better than without banks. It is thus natural to ask at which reserve constraint $\tilde{\alpha}$ welfare is the same in the narrow banking and in the fractional reserve banking system. We get the following proposition:

Proposition 4. *There is a threshold value $\tilde{\alpha} = \frac{\gamma/\beta}{\gamma/\beta+s} > 0.5$ where welfare in the fractional reserve banking economy is identical to the narrow banking economy. If $\alpha > \tilde{\alpha}$ welfare in the narrow banking economy is higher and if $\alpha < \tilde{\alpha}$ welfare in the fractional reserve banking economy is higher.*

In the model without preference shock fractional reserve banking always dominated narrow banking because it allowed interest payments on inside money (see proposition 1). Now, with preference shock, we get a threshold result. Why is this? What is the disadvantage of fractional reserve banking if there is liquidity risk? The answer is that in contrast to narrow banking the insurance against the preference shock is imperfect with fractional reserve banking. We saw that in the narrow banking economy the equilibrium interest rate for cash deposits equals the inflation tax or $1 + i_d = \gamma/\beta$ and thus sellers are fully compensated for the inflation tax on the cash they bring into the period. This is not the case in the fractional reserve banking economy. There we saw that the equilibrium interest rate for cash deposits (with equals the interest on inside money) is below the inflation tax or $1 + i_d < \gamma/\beta$. Thus insurance against the preference shock is imperfect and this is the reason why the narrow banking economy can be better in terms of welfare.

How can we interpret the threshold $\tilde{\alpha}$? Remember that the relative advantage of fractional reserve banking is the provision of interest. This advantage is specially valuable if the inflation tax is high. So the higher the inflation tax the more likely is the fractional reserve system better than the narrow banking system. This is why $\tilde{\alpha}$ increases with the inflation tax. Second, interest payments are negatively related to α. The lower α the higher interest payments on inside money. This increases the relative advantage of fractional reserves. Lastly there is the preference shock or the risk of becoming a seller in the first market s. Clearly perfect insurance against this risk is more valuable if the risk is high (s is high). Thus it makes sense that $\tilde{\alpha}$ decreases in s. The higher this risk, the more likely is the narrow banking system to be better than the fractional reserve banking system. Note that $\tilde{\alpha}$ 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.

At this point one might ask the following question: In the narrow banking economy we saw that interest on deposited cash was exactly equal to the inflation tax or $1 + i_d = \frac{\gamma}{\beta}$. So since agents seem to be fully compensated for the inflation tax why don't we achieve efficiency there? The point is that in the narrow banking economy interest is only paid on the cash deposited in banks and not on the whole stock of outside money. More precisely, interest is paid on the cash *not* used as means of payments. It only applies to a fraction s of the total stock of cash deposited by sellers. 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 sellers (fraction s of the total stock). In the 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.

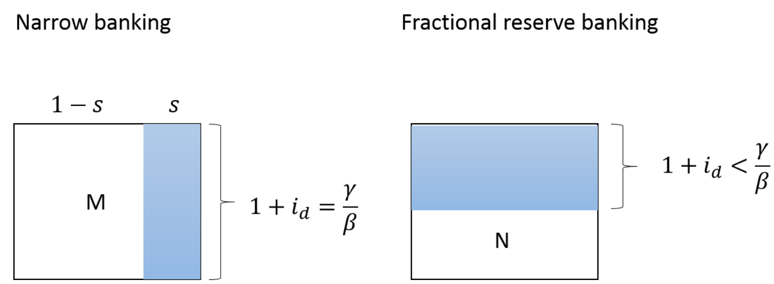


Figure 2: Interest under narrow and fractional reserve banking

Interest in the narrow banking economy is fully targeted to insure agents against the preference shock but the basic inefficiency in the environment - the inflation tax - is not addressed. In the fractional reserve banking economy the basic inefficiency (inflation tax) is partially addressed by the interest on inside money. However, since this interest rate is below the inflation tax, imperfect insurance against the preference shock is reintroduced.

Which system is more efficient if the interest payments per unit of money are the same in both systems? In the narrow banking economy the total amount of interest generated is the total amount of loans $(1-s)l_b$ times the interest payment which in equilibrium equals $i = \gamma/\beta - 1$. In equilibrium total loans equal sM from equation (37). So total interest per unit of money in the narrow banking economy is $s(\gamma/\beta - 1)$. In the fractional reserve banking economy total loans in equilibrium equal $\frac{1-\alpha}{\alpha}M$ from loan market clearing (31) and the loan rate payments i are given by equation (33). Total inside money in equilibrium $N = M + L$ then equals M/α. Thus total interest payments per unit of inside money in the fractional reserve banking economy are

$$\frac{Li}{N} = \frac{1-\alpha}{\alpha} M \frac{\alpha}{M} \frac{(\gamma/\beta - 1)}{(1-\alpha s)} = \frac{(1-\alpha)(\gamma/\beta - 1)}{1-\alpha s}$$

which equals i_d in (33) as it should be. Now we can find a threshold $\hat{\alpha}$ where interest payments per unit of money are the same in both systems. It solves

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

which is

$$\hat{\alpha} = \frac{1-s}{1-s^2}$$

Comparing this threshold with $\tilde{\alpha}$ from proposition 4 where welfare in both regimes is identical we find that $\tilde{\alpha} > \hat{\alpha}$ for $\gamma/\beta > 1$. Thus at $\hat{\alpha}$ where the interest payments per unit of money are the same in both systems welfare is higher in the fractional reserve system than in the narrow banking system. It is more efficient to pay a lower interest rate on the whole stock of money than a higher, fully-compensating interest rate on part of the money stock.

4.5 Two types of debt

So far there was the implicit assumption that fractional reserve banks can only issue inside money as liabilities which they have to back with a fraction α in outside money. Now suppose they also have the possibility to issue non-monetary debt (debt not used as money in the first market at interest i'_d) which they don't need to back with outside money. The only agents who could hold this debt are the sellers who have idle cash after the preference shock but don't need money in the first market. In fact, if $i'_d > i_d$ they would prefer depositing their cash in the bank against non-monetary debt than against acquiring inside money. Suppose this holds so there is a demand for this debt and denote d' as cash deposits against non-monetary debt (per buyer). The bank problem now reads:

$$\max_{l,d,d'} = d + d' + l(1+i) - (l+d)(1+i_d) - d'(1+i'_d)$$
$$s.t. \quad \alpha(l+d) \leq d + d'$$

As can be seen from the constraint non-monetary debt has the advantage for the bank 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 $(1-\alpha)/\alpha$. If for example $\alpha = 0.5$ a marginal cash deposit against non-monetary debt increases loan capacity by 2 but a cash deposit against inside money only by 1. The disadvantage of non-monetary debt are the higher costs since $i'_d > i_d$.

As before if $i > i_d$ and $i_d, i'_d > 0$ the constraint must bind and we can rewrite the objective function:

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

Thus if both types of debt are used and we apply free entry (zero profits) we get again that the interest rate on inside money is a weighted average of the return on outside money and loans (5):

$$1 + i_d = \alpha + (1-\alpha)(1+i)$$

And the interest rate on non-monetary debt must satisfy (the last step follows from applying (5):

$$1 + i'_d = 1 + \frac{1+i}{\alpha} - \frac{1+i_d}{\alpha} = 1 + i > 1 + i_d \tag{40}$$

Thus if the interest on inside money is this weighted average the bank can pay a higher interest rate equal to the loan rate on non-monetary debt and there is no spread. Thus we see $i'_d > i_d$ must hold. This interest rate structure has interesting implications for the model. We know the seller will deposit against non-monetary debt if $i'_d > i_d$. Thus the marginal value of outside money of a seller (29) is now

$$V'_s(m) = \phi(1 + i'_d) \tag{41}$$

and optimal outside money holdings (30) solve:

$$\phi = \beta\phi_{+1}[(1-s)(1+i_{+1}) + s(1+i'_{d+1})]. \tag{42}$$

Thus in equilibrium we get the equivalent to (32)as:

$$\frac{\gamma}{\beta} = (1-s)(1+i) + s(1+i_d) = 1 + i = 1 + i'_d \tag{43}$$

Thus as in the basic model without preference shock in section 3 the loan rate is equal to the inflation tax and independent of α. This is in contrast to the model with only inside money. Since (5) also holds we get the same equilibrium interest rate on inside money (20) and the same equilibrium allocation (22) as in the model without preference shock:

$$\frac{u'(q_b)}{c'(\frac{1-s}{s}q_b)} = \frac{\gamma/\beta}{(1-\alpha)\gamma/\beta + \alpha} \tag{44}$$

From this we can draw the following conclusions:

Proposition 5. *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 (44)and equals the allocation without preference shock in proposition 1.

ii) the allocation is better than in an economy with only inside money in proposition 2.

iii) the allocation is better than the narrow banking allocation in proposition 3 and converges to this allocation as $\alpha \to 1$.

With two types of debt the threshold result from the economy with only inside money in proposition 2 vanishes and fractional reserve banking dominates narrow banking as in the basic model in section 3. The intuition is the following: With only inside money narrow banking had the advantage of a better insurance against liquidity risk because sellers were perfectly compensated for the inflation tax on the cash they acquired whereas under fractional reserve banking they were not. This advantage disappears now because non-monetary debt also perfectly compensates the sellers against the inflation tax. It pays an interest rate equal to the inflation tax in equilibrium. With two types of debt also fractional reserve banking offers perfect insurance against liquidity risk.

Finally a few remarks on the interpretation of the model with $\alpha = 1$ as a narrow banking system. If we write the bank problem with $\alpha = 1$ it becomes:

$$\max_{l,d,d'} = d + d' + l(1+i) - (l+d)(1+i_d) - d'(1+i'_d)$$
$$s.t. \quad l \leq d'$$

One can see from the objective function that banks only accept cash deposits for inside money d if they have to pay zero interest, i.e. if $i_d = 0$. Otherwise they would set $d = 0$. One can also see from the constraint that the capacity to make loans in inside money is determined by the cash deposits against non-monetary debt d'. These deposits increase outside money holdings of banks. Thus even if banks have to back inside money one-to-one with outside money they can issue inside money over loans if they get additional cash deposits. Now suppose buyers prefer inside money over outside money even if $i_d = 0$ and both types of money are exact substitutes. In this case there would still be a demand for cash deposits against inside money from buyers after the preference shock. Going with this scenario (and a binding constraint, i.e. $l = d'$) the zero profit condition again gives us $i = i'_d$. Thus a narrow bank with these two types of debt has outside money holdings of $d' + d$ which fully back the issued inside money holdings $d + l$ since $l = d'$. And she has loans in inside money l financed by non-monetary debt d'. Inside money pays zero interest but non-monetary debt is remunerated at the same interest rate as inside money loans. Thus we are at the same allocation as in the narrow banking economy in section 4.3.

5 conclusion

The specific focus of this paper points to various extensions which could be addressed in further work. First: the models presented here neglects the fundamental role of banks in capital allocation and investment as it is essentially a model of efficient liquidity provision and allocation. Second: it could be interesting to quantify the welfare gains from inside money creation. Some quantitative estimates from New Monetarist models are available. They could be complemented with a quantitative estimate of this model. Third: The analysis is embedded in an environment of perfect competition. This aligns the private interest of banks with social welfare. But e.g. the discussion of the desirability of private seignorage by banks is impossible in a model where banks make zero profits. Fourth: the analysis abstracts from the fragility of fractional reserve banking systems. But a complete comparison between the two banking systems should include it.

References

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

Ricardo Cavalcanti, Andres Erosa, and Ted Temzelides. Private money and reserve management in a random-matching model. *Journal of Political Economy*, 107(5):929–945, 1999.

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

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

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

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.

Thomas J. Sargent and Neil Wallace. The real-bills doctrine versus the quantity theory: A reconsideration. *Journal of Political Economy*, 90(6):1212–1236, 1982.

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 only outside money

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:

$$\frac{u'(q_b)}{c'(q_s)} = \frac{\gamma}{\beta} \tag{45}$$

For a positive inflation tax ($\gamma > \beta$) the allocation is inefficient and the inefficiency increases in the inflation tax. To see this note that under the assumptions made on preferences, especially the concavity of $u(q)$, the LHS of 45 $\frac{u'(q)}{c'(q)}$ decreases in q:

$$\frac{d(\frac{u'(q)}{c'(q)})}{dq} = \frac{u''(q)c'(q) - u'(q)c''(q)}{c'(q)^2} < 0 \tag{46}$$

So if $\gamma/\beta > 1$ it must be that $q < q^*$ given by (1).

A.2 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 outside money and that this need is proportional to the inside money used (α). This was a shortcut because the concrete clearing/settlement process was not modelled. In this section I build a model building on Williamson [2012] where the outside money holdings of banks serve a deeper purpose. As in the basic model buyers acquire inside money in the second market. Afterwards buyers either go to a *non-monitored* meeting with probability π where they can only use outside money. Or with the inverse probability they go to a *monitored* meeting where they can use inside money. As before 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 (at the beginning of the first market) buyers can redeem inside money into outside money at interest rate i_{d1}. Inside money which is not redeemed early will be redeemed in the following second market at the interest rate i_{d2} as before. Banks thus hold outside money because some buyers need it in the first market and want to redeem their inside money holdings early.

Introducing early redemptions gives banks more flexibility in the interest rates they can pay in the second market, depending on how many buyers redeem early at at what conditions. The basic message from the sections before is repeated and refined in this section. In general, more inside money creation is beneficial also in this framework because it increases the return on inside money.

As in Williamson [2012] I assume take-it-or-leave-it offers by buyers, a unit mass of buyers and sellers and linear disutility of working in the first market, $c(q) = q$ for sellers. I also assume $-q\frac{u''(q)}{u'(q)} < 1$ and more specifically $u(q) = \frac{q^{1-\sigma}}{1-\sigma}$ with $\sigma < 1$). Let the amount consumed in a non-monitored meeting be q^c (for "cash") and q in a monitored meeting. n^c (n') are the amounts of inside money a buyer does not redeem in a non-monitored (monitored) meeting. The problem of a representative buyer can then be written as:

$$\begin{aligned}\max_{d,l,n^c\geq 0,n'\geq 0} \quad & -\phi d + \pi[u(\beta\phi_{+1}(d+l-n^c)(1+i_{d1})) + \beta\phi_{+1}n^c(1+i_{d2})] \\ & (1-\pi)[u(\beta\phi_{+1}(d+l-n')(1+i_{d2})) + \beta\phi_{+1}n'(1+i_{d2})] \\ & -\beta\phi_{+1}l(1+i)\end{aligned} \tag{47}$$

A buyer works to acquire outside money and deposit it in the bank (d) or she might borrow inside money (l). In a non-monitored meeting the buyer redeems $d+l-n^c$ units of inside money for cash at rate $1+i_{d1}$. The value of this outside money is $\beta\phi_{+1}(d+l-n^c)(1+i_{d1})$ 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. In the case of a monitored meeting the buyer uses inside money to pay the seller where the marginal value in the next second market is $\beta\phi_{+1}(1+i_{d2})$. 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:

$$u'(\tilde{q}^c) = \frac{1+i_{d2}}{1+i_{d1}} \quad , \quad u'(q^*) = 1 \tag{48}$$

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 marginal costs $1+i_{d2} = 1+i = \frac{\phi}{\beta\phi_{+1}}$ and the buyer doesn't use all his inside money in both types of meetings ($n^c, n' > 0$). The amount of real inside money holdings $\beta\phi_{+1}n$ is not determined and must lie in $(\frac{q^*}{1+i_{d2}}, \infty)$.[24] Defining $r = (n - n^c)/n$ 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 $r = \frac{\tilde{q}^c}{\beta\phi_{+1}n(1+i_{d1})} \in \big(\frac{\tilde{q}^c(1+i_{d2})}{(1+i_{d1})q^*}, 0\big)$.

b) medium return on inside money ($i_{d1}, \tilde{i}_{d2} < i_{d2} < i$): In this case inside money is costly to hold ($i_{d2} < i$) but the return is still high enough that buyers are

[24] To be unconstrained in the monitored meeting a buyer needs $q^*/(1+i_{d2})$ units of real inside money. To be unconstrained in the non-monitored meeting the buyer needs $\tilde{q}^c/(1+i_{d1})$. With $\sigma < 1$ we have $\tilde{q}^c/(1+i_{d1}) < q^*/(1+i_{d2})$ and thus buyers need less real inside money to be unconstrained in the non-monitored meeting.

unconstrained in the non-monitored meetings, i.e. $q^c = \tilde{q}^c$ and $n^c > 0$ but constrained in the monitored ones ($n' = 0$). The consumption levels are:

$$u'(\tilde{q}^c) = \frac{1+i_{d2}}{1+i_{d1}} \quad , \quad u'(q) = \frac{1+i-\pi(1+i_{d2})}{(1-\pi)(1+i_{d2})} \tag{49}$$

and real inside money holdings solve $\beta\phi_{+1}n = q/(1+i_{d2})$. The redemption rate is $r = \frac{\tilde{q}^c(1+i_{d2})}{q(1+i_{d1})}$. Note that real inside money holdings increase in the return on inside money i_{d2}. Thus we can find a lower bound on i_{d2} where real inside money holdings just equal $\tilde{q}^c$ when all inside money is redeemed i.e. when $r = 1$. This threshold solves:

$$\beta\phi_{+1}n(1+i_{d1}) = \frac{q}{1+\tilde{i}_{d2}}(1+i_{d1}) = \tilde{q} \tag{50}$$

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 ($n^c = n' = 0$). The consumption levels and real inside money holdings solve:

$$\begin{aligned} \pi u'(q^c)(1+i_{d1}) + (1-\pi)u'(q)(1+i_{d2}) &= 1+i \\ q^c = \beta\phi_{+1}n(1+i_{d1}) \quad , \quad q &= \beta\phi_{+1}n(1+i_{d2}) \end{aligned} \tag{51}$$

which yields:

$$\beta\phi_{+1}n = \Big(\frac{\pi(1+i_{d1})^{1-\sigma} + (1-\pi)(1+i_{d2})^{1-\sigma}}{1+i}\Big)^{1/\sigma} \tag{52}$$

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

$$u'(q) = \frac{1+i}{1+i_d} \tag{53}$$

with real inside money holdings $\beta\phi_{+1}n = q/(1+i_d)$.

In any of these cases the indifference condition between the two ways of acquiring inside money (borrowing and depositing cash, (9)) must hold as in the basic model:

$$\frac{\phi}{\phi_{+1}}\frac{1}{\beta} = 1 + i_{+1} \tag{54}$$

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

$$0 \le i_{d1} \le i_{d2} \le i \tag{55}$$

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

bank problem: The bank maximizes total cash profits taking the redemption rate r as given. Total redemptions before the first market are given by $\pi rn(1+i_{d1})$[25] 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 $\bar{l}$.[26] 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 influence the bank problem in two ways (compare it with2). It decreases outside money holdings and it decreases the outstanding inside money of banks. Thus it both decreases 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:

$$\bar{l} = \frac{1-\pi r(1+i_{d1})}{\pi r(1+i_{d1})} d \tag{56}$$

and the zero profit condition implies:

$$1 + i_{d2} = \frac{1-\pi r(1+i_{d1})}{1-\pi r}(1+i) \tag{57}$$

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 (54) must hold, this can never be an equilibrium because the inflation tax is above

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

[26] 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 loans/private debt in my model). There is also a quantity constraint which steers the aggregate issuance of government bonds, $\delta = \frac{M}{M+B}$. Note that this constraint is very similar to the $\alpha = \frac{M}{M+L}$ constraint in the basic model which steers the composition of outside and inside money.

1. If the loan rate lies below the inflation tax, agents would not be willing to deposit outside money in the bank. 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

$$(i - \pi r i_{d1} + (1 - \pi r) i_{d2})\bar{l} = (\pi r i_{d1} + (1 - \pi r) i_{d2}) d \tag{58}$$

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), (58) and $d > \pi r n(1 + i_{d1})$ which implies

$$1 + i_{d2} < \frac{1 - \pi r(1 + i_{d1})}{1 - \pi r}(1 + i) \tag{59}$$

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 (54) as in the normal model. Also market clearing for outside money implies $d = M$. Thus the aggregate ratio of outside to inside money under a binding borrowing constraint is fixed and we can define this ratio as $\bar{\alpha}$[27]

$$\bar{\alpha} = \frac{M}{M + \bar{l}}$$

We can rewrite equilibrium interest rates without excess reserves using (56) and (57) as

$$\begin{aligned} 1 + i_{d1} &= \frac{\bar{\alpha}}{\pi r} \\ 1 + i_{d2} &= \frac{(1 - \bar{\alpha})(1 + i)}{1 - \pi r} \end{aligned} \tag{60}$$

With excess reserves using (58) we get:

$$\pi r(1 + i_{d1}) + (1 - \pi r)(1 + i_{d2}) = \bar{\alpha} + (1 - \bar{\alpha})(1 + i) \tag{61}$$

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 ($i_{d2} = i$). Therefore it must be from (60) that $\bar{\alpha} = \pi r$ and $i_{d1} = 0$, i.e. the interest rate for early redemptions is zero. Given these interest rates equilibrium consumption in both types of meetings solves:

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

[27]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.

Since $\tilde{q}^c = \beta\phi_{+1} r n = \frac{\beta\phi_{+1}\bar{\alpha} n}{\pi} = \frac{\beta\phi_{+1} M}{\pi}$ real outside money holdings are given by $\beta\phi_{+1} M = \pi\tilde{q}^c$. We also know real inside money holdings $\beta\phi_{+1} n$ must be at least $\frac{q^*}{\gamma/\beta}$. Thus for existence of this equilibrium $\bar{\alpha}$ must solve

$$\beta\phi_{+1} n = \frac{\beta\phi_{+1} M}{\bar{\alpha}} \geq \frac{q^*}{\gamma/\beta}$$
$$\bar{\alpha} \leq \frac{\pi\tilde{q}^c(\gamma/\beta)}{q^*} = \alpha^* < \pi \tag{62}$$

which exists for $\bar{\alpha} \in (0, \alpha^*)$.

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. 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 (62). 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 redeem 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 $(q, \tilde{q}^c, i_{d1}, i_{d2}, r, \beta\phi_{+1} n)$:

$$u'(\tilde{q}^c) = \frac{1 + i_{d2}}{1 + i_{d1}} \quad , \quad u'(q) = \frac{1 + i - \pi(1 + i_{d2})}{(1 - \pi)(1 + i_{d2})}, \tag{63}$$

(60), $\beta\phi_{+1} n = q/(1 + i_{d2})$ and $r = \frac{\tilde{q}^c(1+i_{d2})}{q(1+i_{d1})}$. We can reduce them to one equation which implicitly defines i_{d2}:

$$(1 - \pi)(1 + i_{d2})^{2-\sigma} - (1 - \pi)(1 - \alpha)(1 + i)(1 + i_{d2})^{1-\sigma}$$
$$+ \pi^{\sigma+1}\alpha^{1-\sigma}(1 + i_{d2}) - \pi^{\sigma}\alpha^{1-\sigma}(1 + i) = 0 \tag{64}$$

We know the equilibrium should exist for $\tilde{i}_{d2} < i$ where $\tilde{i}_{d2}$ is pinned down by (50) with a redemption rate of 1. We can thus define a threshold $\tilde{\alpha}$ which solves (50) for interest rates satisfying (60) with $r = 1$. Rewriting (50) we get:

$$\left(\frac{1 + i_{d2}}{1 + i_{d1}}\right)^{1-\sigma} = \frac{1 + i - \pi(1 + i_{d2})}{(1 - \pi)(1 + i_{d2})} \tag{65}$$

Suppose $\tilde{\alpha}$ is π. From (60) we would then get $i_{d2} = i$ and $i_{d1} = 0$. Clearly this violates (65). We need $\tilde{i}_{d2} < i$ and ergo $\tilde{\alpha} > \pi$. Now suppose $\tilde{\alpha} = \underline{\alpha}$ where

$$\underline{\alpha} = \frac{\pi(\gamma/\beta)}{1 - \pi + \pi(\gamma/\beta)} \tag{66}$$

is the value where both interest rates are equal ($i_{d2} = i_{d1} = i_d$), according to (60) with $r = 1$. Also then (65) 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 (..) we must have $\tilde{\alpha} < \alpha^*$ and the interval $(\alpha^*, \tilde{\alpha})$ is non-empty.

(64) is consistent with these two boundary conditions. Also, using the implicit function theorem it can be shown that (64) implies $d(1 + i_{d2})/d\alpha < 0$ for $i_{d2} \in (\tilde{i}_{d2}, i)$. As $\bar{\alpha}$ decreases from $\tilde{\alpha}$ to α^* i_{d2} increases from $\tilde{i}_{d2}$ to i. Given this we can deduce the behaviour of the other variables in equilibrium. At $\tilde{\alpha}$ $i_{d1} > 0$ and at α^* $i_{d1} = 0$ thus the interest rate for early redemptions decreases. The redemption rate r is 1 at $\tilde{\alpha}$ and $\frac{\tilde{q}^c(\gamma/\beta)}{q^*} < 1$ at α^* thus it also decreases. Consumption in the monitored meeting increases in i_{d2} from (63) and thus increases as $\bar{\alpha}$ goes to α^*. On the other hand consumption in the non-monitored meeting decreases as $\bar{\alpha}$ decreases to α^*.

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 (47):

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

Using (57) to substitute for $1 + i_{d1}$ and applying the envelope theorem:

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

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.[28]

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 (51) and (60) 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 $\bar{\alpha} \in (\underline{\alpha}, \tilde{\alpha})$.

Using (52) one can show that $\frac{d(\beta\phi_{+1}n)}{d\alpha} < 0$ 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

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

the monitored meeting increase. The positive effects on inside money holdings from the increase in i_{d1} are overcompensated by the negative effects over 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[29] 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 (57) to substitute for $1+i_{d1}$ and applying the envelope theorem we see that:

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

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]. In his model monitored buyers (and expected) welfare also increase with more bonds issued (a lower $\bar{\alpha}$). The consumption of uyers in non-monitored meetings however, is unaffected from issuing more bonds. Bonds are only used in monitored meetings and increasing the quantity (and the return) of bonds by decreasing $\bar{\alpha}$ only benefits monitored buyers. There are no spillover effects.

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 (60) 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 (61):

$$1 + i_d = \bar{\alpha} + (1-\bar{\alpha})(1+i) \tag{67}$$

Thus this equilibrium solves (53) and (67) and $\beta\phi_{+1}n = q/(1+i_d)$ in $\bar{\alpha} \in (\underline{\alpha}, 1)$. Note that (67) is exactly the same expression we got in the basic model (5). 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 (22) (with a linear disutility of working for sellers, i.e. $c'(q_s) = 1$)

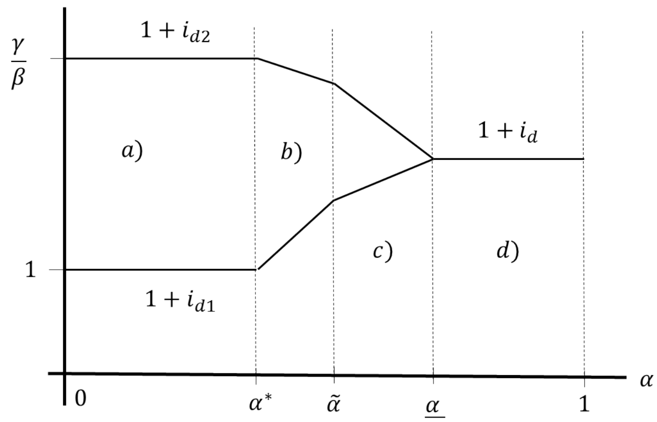
$$u'(q) = \frac{1+i}{1+i_d} = \frac{\gamma/\beta}{(1-\alpha)\gamma/\beta + \alpha} \tag{68}$$

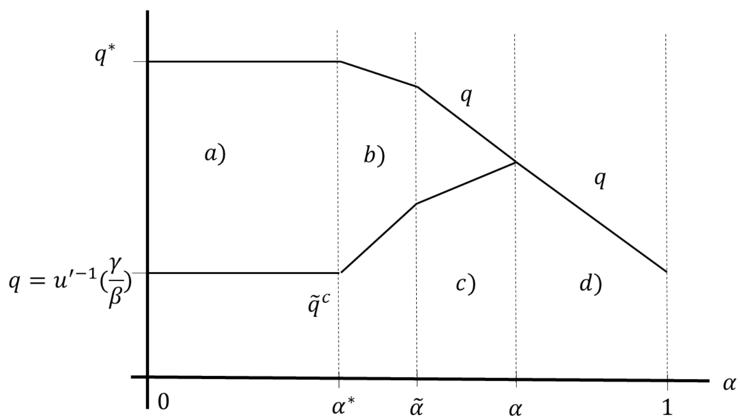
[29]Concretely if $\gamma/\beta < \frac{\pi+(1-\pi)\sigma}{\pi(1-\sigma)}$.

The welfare implications are also identical to the basic model. The allocation is non-neutral with respect to $\bar{\alpha}$. The higher $\bar{\alpha}$ (the less inside money) the lower welfare.[30] 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.

[30] 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 (54). 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.





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

Let p^c denote the price of the first-market-consumption good q in outside money. For outside money to be accepted besides inside money, agents must be indifferent between using inside or outside money. For sellers, indifference means they could produce the same amount q_s paid with either means of payment. If they work for cash their marginal utility is $p^c\phi$ and if they work for inside money $p\phi(1+i_d)$ so to be indifferent the ratio of prices must be

$$\frac{p^c}{p} = 1 + i_d$$

which is > 1 if $i_d > 0$. Sellers need to be compensated with a higher price for sales against cash. Indifference for buyers means they can consume the same amount q_b in the first market with outside or inside money for the same disutility of working to acquire it in the second market before. So they would acquire m units of outside money such that:

$$m/p^c = n/p = q_b \tag{69}$$

But since the price of outside money in the first market is higher ($p^c > p$) when sellers are indifferent, this implies $m > n$ i.e. buyers need to acquire more outside money than outside money to consume the same amount of goods. However, both means of payment have the same marginal costs ϕ in the second market (wlog we can assume a buyer would only deposit cash to acquire inside money). Thus the costs of acquiring m units of outside money are strictly higher than the costs of acquiring n units of inside money and a buyer would strictly prefer to use inside money over outside money in the first market.

A.4 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 (26) 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 $\tilde{q}_b$ solves:

$$\frac{u'(q_b)}{c'(\frac{1-s}{s}q_b)} = \frac{\gamma/\beta - s}{1-s}. \tag{70}$$

We see the RHS of equation (70) is equal to the RHS of 45 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 45 and the inefficiency increases in s.

Appendix B

B.1 proof of proposition 1

Proof. I first show that the conjectures made to derive (22) are indeed correct in equilibrium for $\alpha \in (0, 1)$ and $\gamma > \beta$. The conjectures were:

1. $i > i_d > 0$ and the bank reserve constraint binds
2. sellers don't acquire inside money
3. buyers are liquidity constrained

The first conjecture is easily verified by looking at the expressions for optimal interest rates in equilibrium, (20). For an interior α and a positive inflation tax $i > i_d > 0$ holds in equilibrium and thus the bank constraint binds.

The second conjecture: Since for a seller inside money has no liquidity value and his preferences in the second market are quasi-linear he will not acquire inside money if the marginal costs ϕ are higher than the discounted marginal benefits $\beta\phi_{+1}(1 + i_{d+1})$ In equilibrium this is satisfied if the interest on inside money is below the inflation tax, $1 + i_d < \gamma/\beta$ and this condition also holds for $\alpha \in (0, 1)$ and $\gamma > \beta$ according to (20). Also, a seller will not acquire inside money by borrowing if the marginal costs $\phi_{+1}(1 + i_{+1})$ are higher than the marginal benefits $\phi_{+1}(1 + i_{d+1})$ or in other words if $i > i_d$. Thus the conditions on interest rates shown to be true in equilibrium in conjecture 1 also imply conjecture 2.

To verify the third conjecture we see from optimal consumption of the buyer (15) that a buyer is liquidity constrained if $q_b < q^*$. In equilibrium this is equivalent to saying that the q_b given by (22) is inefficient (below q^*) or that the RHS of (22) is above 1 (see A.1). Since the RHS of (22) can also be written as the ratio between (gross) loan and inside money rate $\frac{u'(q_b)}{c'(q_s)} = \frac{1+i}{1+i_d}$ $q_b < q^*$ holds if $i > i_d$ which was already shown to be true in conjecture 1. This also establishes the inefficiency of the allocation statet in ii).

Next I show that given that the buyer is liquidity constrained in equilibrium, i.e. $q_b = n/p$, the problem in (6) is concave in cash deposits d and loans l and thus the first order conditions for optimal inside money holdings (7) uniquely maximize the problem in (6). Problem (6) writes for a liquidity constrained buyer:

$$W(0, l_{-1}) = \max_{x,d,l} U(x) - l_{-1}\phi(1 + i) - x - \phi d + \phi\tau + \beta(u(\frac{d + l}{p_{+1}}) - l\phi_{+1}(1 + i_{+1}))$$

which is concave in d and l by the concavity of $u(q)$. Thus the solution in (22) is unique.

For the comparative statics of the equilibrium allocation stated in *ii*) I refer to A.1 where I showed that the closer the RHS of (22) to 1 the better for welfare. Thus since $\frac{\partial(\frac{u'(q_b)}{c'(q_s)})}{\partial\gamma/\beta} > 0$ and $\frac{\partial(\frac{u'(q_b)}{c'(q_s)})}{\partial\alpha} > 0$ the RHS of (22) increases both in the inflation tax an in α which decreases welfare.

To see that the allocation in (22) is better than the allocation without inside money creation (45) stated in *iii*) note that the RHS of (22) can be written as $\gamma/\beta\frac{1}{1+i}$ which is below γ/β (the RHS of (45) if interest on inside money is positive which is true in this economy. The results *iv*) and *v*) are simply obtained by sticking in the limit values of α $(0, 1)$ into the RHS of 22.

■

B.2 proof of proposition 2

Proof. The only additional conjecture in the model from section 4 compared to section 3 is that if we are in an equilibrium where buyers borrow it is optimal for them to borrow. To verify this it must be that the marginal utility if buyers don't borrow $u'(m/p)1/p$ is bigger than the marginal costs of borrowing $\phi(1+i)$ or:

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

which holds if $l > 0$. Thus it is optimal for buyers to borrow additional inside money in this equilibrium. The other conjectures are the same as in section 3 and the proof that (35) uniquely solves the model in section 4 follows the same steps as the proof for the model without preference shock above.

ii): Since $i > i_d$ follows also here that the RHS of (35) is > 1 and thus $q < q^*$. The only new element in the comparative statics is the effect of s. Since $\frac{\partial(\frac{u'(q_b)}{c'(q_s)})}{\partial s} > 0$ the RHS of (35) increases in s which decreases welfare by the same logic as above.

iii): Note that as s goes to zero the allocation given by (35) converges to the allocation without preference shock (22). And since increasing s decreases welfare the allocation in the basic model is strictly better.

iv): Compare the RHS of (70) with the RHS of (35):

$$\frac{\gamma/\beta - s}{1-s} \cdots \frac{\gamma/\beta - \alpha s}{(1-\alpha)\gamma/\beta + \alpha(1-s)}$$
$$\leftrightarrow$$
$$c - s > 1 - s$$

to see that the former is bigger and thus welfare is strictly higher in the fractional reserve banking economy than in the economy without banks. Also see that both RHS are equal if $\alpha = 1$. To get the result from *v*) stick in $\alpha = 0$ in the RHS of (35) and see that it equals 1 which is equivalent to the FB-allocation in (1).

■

B.3 proof of proposition 3

Proof. See the proof of proposition 1 in BCW.

iv): The allocation with outside money and preference shock without banks is given by (A.4):

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

T show that this allocation is worse than the allocation in the narrow banking economy we need:

$$\frac{u'(q_b)}{c'(\frac{1-s}{s}q_b)} = \frac{\gamma/\beta - s}{1-s} > \frac{\gamma}{\beta} = \frac{u'(q_b^{NB})}{c'(\frac{1-s}{s}q_b^{NB})}$$

Which is true for any $s > 0$.

■

B.4 proof of proposition 4

Proof. Finding $\tilde{\alpha}$ requires equalizing the right-hand sides of (39) and (35) thus solving

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

which yields

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

■

B.5 proof of proposition 5

All statements in the proposition have already been proved elsewhere. Statement *ii*) is equivalent to statement *iii*) in proposition 2.