5-18-2019

# Monetary Equilibrium and the Cost of Banking Activity

Paola Boel
*Sveriges Riksbank*

Gabriele Camera
*Chapman University*, camera@chapman.edu

# Monetary Equilibrium and the Cost of Banking Activity†

Paola Boel
Sveriges Riksbank

Gabriele Camera
Chapman University
University of Bologna

May 18, 2019

**Abstract**

We investigate the effects of banks' operating costs on allocations and welfare in a low interest rate environment. We introduce an explicit production function for banks in a microfounded model where banks employ labor resources, hired on a competitive market, to run their operations. In equilibrium, this generates a spread between interest rates on loans and deposits, which naturally reflects the underlying monetary policy and the efficiency of financial intermediation. In a deflation or low inflation environment, equilibrium deposits yield zero returns. Hence, banks end up soaking up labor resources to offer deposits that do not outperform idle balances, thus reducing aggregate efficiency.

Keywords: banks; frictions; matching.
JEL: C70, D40, E30, J30

## 1 Introduction

Banking is an expensive activity because the operating costs due to branch leases and employee compensation are non trivial. In the United States, for example, financial activities employed almost 8.5 million workers and accounted for 7.5% of GDP in 2017.[1] The average European country spends about one percent of GDP to conduct retail transactions, and half of these costs are incurred by banks and infrastructures (Schmiedel et al., 2012). When interest rates are low, these operating costs naturally compress banks' profit margins and ability to generate loans. Moreover, core deposits are likely to decrease in such an environment, as documented by Calomiris and Nissim (2014) for the

†We thank an anonymous referee, and seminar participants at the Bank of Canada, Sveriges Riksbank, Loyola University Chicago and Indiana University. The views expressed are those of the individual authors and do not necessarily reflect official positions of Sveriges Riksbank.

[1]Data for employment are from the *Bureau of Labor Statistics*, as measured by Employment in Financial Activities. Data for the share of GDP are from the St. Louis *FRED Database*, as measured by Value Added by Private Industries: Finance and Insurance.

United States in the last decade. In light of these facts, several natural questions arise: Do banks retain their welfare-improving role at low interest rates when we account for financial intermediation costs? If they do not, what determines the specific interest-rate threshold at which banks cease to exercise their welfare-enhancing role? And what are the quantitative consequences in terms of monetary policy?

We investigate these issues in a microfounded model of banking where money has an explicit role. The presence of money here is relevant because we want agents to have the option of holding cash instead of keeping funds in the bank if deposit rates become too low. Our study thus builds on Berentsen, Camera and Waller (2007) (henceforth BCW), who demonstrated from first principles how banks improve social welfare by reallocating liquidity from those with idle balances to those who are liquidity constrained. Their central result is that banks improve welfare for *any* positive nominal interest rate, even for policies arbitrarily close to the zero bound. This is because introducing banks always raises the return from holding idle balances. Such result emerges provided that the process of financial intermediation requires neither labor nor capital inputs, that is, banks have no operating costs. Though this assumption has benefits in terms of theoretical simplicity, it abstracts away from the important empirical feature that banks incur substantial costs in order to conduct their activities.

Here, we study the existence of an equilibrium when intermediation costs are explicit: intermediating a loan requires labor resources. Our model generates a well-defined bid-ask spread for interest rates which endogenously responds in natural ways to monetary policy and the efficiency of the intermediation technology. This in turn implies that financial intermediation costs are quantitatively important. If the spread between rates on loans, deposits and bonds is taken to measure market liquidity, then market liquidity is affected by monetary policy, whereas the spread is always zero when financial intermediation is costless.

A main result of our model is that a banking equilibrium does not always exist close to the zero bound. It instead generally exists only for nominal interest rates bounded sufficiently away from zero. Intuitively, banks' revenues and deposit rates shrink as rates fall, and at some point paying interests on deposits is no longer profitable. Eventually, lending will also cease to be profitable if banks are sufficiently inefficient. If so, then a deflation that brings the economy close to the Friedman rule is incompatible with the existence of banks. Supporting a banking equilibrium when interest rates are close to zero requires improvements in the efficiency with which banks transform liquidity into loans or else banks will collapse. This is in contrast with the prediction in BCW where interests on deposits are invariably positive away from the Friedman rule because banking soaks up no resources.

But should we care about the efficiency of financial intermediation? And is an economy without banks worse off than one where banks are active? Quantitatively, we find that intermediation efficiency and welfare cost of inflation exhibit a negative relation. Intuitively, greater efficiency in banking lessens the adverse effect of inflation on market liquidity. However, banks absorb labor resources away from other productive activities, which—if the financial sector is large and inefficient[2]—has an appreciable negative general equilibrium effect on wages, hence on prices. As a result, banks may reduce overall efficiency in a deflation, away from the Friedman rule, and even at small positive inflation rates—when deposit rates are zero. In this sense, the analysis shows that there can be "too much" finance for sufficiently low inflation. In this scenario paying interests on reserves—financing it through lump sum taxes—can correct the problem, as it allows banks to raise deposit rates.

Our study is related to existing microfounded models of banking in which intermediation generates *exogenously* fixed costs. In Bencivenga and Camera (2011) banks are coalitions of agents that accumulate goods which can be then transformed into capital or used to pay withdrawal costs generated by buyers who make withdrawals. Our model is different along two key dimensions. First, in Bencivenga and Camera (2011) the operating cost per unit of withdrawal is exogenous, and thus banks face operating costs only if there are some withdrawals. By contrast, in our model banks' operating costs depend on prevailing wages and the amount of intermediation they undertake. Second, in Bencivenga and Camera (2011) banks are somewhat unrealistic good-storage technologies whose operating costs are in terms of real resources. In that model banks must store goods in order to cover the withdrawal costs which, therefore, dissipate real resources and reduce the amount of goods that can be transformed into capital. By contrast, we consider a more realistic model in which banks do not store goods, and their operating costs

[2]Philippon and Reshef (2013) find that the finance industry's share of gross domestic product (GDP) has increased in the United States since the 1980s. However, Philippon (2015) finds no evidence that the finance industry has become more efficient in recent years, despite organizational changes and advances in information technology.

take the form of workers' wages. Another related paper is Chiu and Meh (2011), which also studies the relationship between inflation and costly banking. In their environment, however, banks face no operating costs. Instead, entrepreneurs suffer a fixed exogenous disutility cost when borrowing money from banks. This implies financial intermediation is welfare improving only at high inflation, that is when the inflation rate is sufficiently large relative to the fixed cost. Thus, contrary to our study, banks' efficiency plays no role in mitigating such costs.

There are also models of costly financial intermediation where banks intermediate real assets but there is no role for money (e.g., see Antunes et al., 2013, and references therein). By contrast, we work with a model where money has an explicit role, and banking costs are due to an explicit labor-based production technology for loans, so labor costs suffered by banks vary with the intermediation level. We are also related to microfounded studies that have investigated the welfare implications of the availability of credit in a BCW environment, albeit without explicitly modeling labor costs for banks. Among them, Chiu et al. (2018) find that credit arrangements can be welfare reducing in competitive markets. That is because high consumption by credit users drives up the price level, thus reducing consumption by money users who are subject to a binding liquidity constraint.

We proceed as follows. Section 2 describes the model, Section 3 discusses existence of an equilibrium, Section 4 presents the main theoretical results, Section 5 discusses the quantitative performance of the model and the last Section concludes.

# 2 The environment

The model builds on BCW (2007). Time is discrete, the horizon is infinite and there is a large population of infinitely-lived agents who produce and consume perishable goods and have time-separable preferences. In each period agents may visit two sequential anonymous markets, denoted by 1 and 2. Markets differ in terms of economic activities and preferences and in each market, workers can access a linear goods' production technology that transforms one unit of labor into one unit of consumption goods. In market 1, agents face an idiosyncratic trading risk. An agent can either consume or work, two mutually exclusive individual states. She can consume with probability $\alpha$ and can work with probability $1 - \alpha$. Since consumers will want to buy consumption goods, we call them buyers. Buyers get utility $u(q)$ from $q > 0$ consumption, with $u' > 0$, $u'' < 0$, $u'(0) = \infty$, and $u'(\infty) = 0$. Supplying $N \geq 0$ labor generates disutility $c(N)$, with $c' > 0$, $c'' \geq 0$ and $c(0) = 0$. In market 2 every agent can consume and produce using a backyard technology, supplying $Y$ labor generates $Y$ disutility, consuming $X$ goods generates utility $U(X)$, with $U' > 0$, $U'' \leq 0$, $U'(0) = \infty$ and $U'(\infty) = 0$. We let $U'(X^*) = 1$. The discount factor across periods is $\beta \in (0, 1)$. We retain the same informational structure as in BCW, which gives rise to a role for monetary exchange and banking activity.

A central bank exists that controls the supply of fiat money. Let $M_t$ denote the per capita money stock at the start of a generic period $t$. The gross growth rate of the money stock across two periods is $\mu \geq \beta$. Agents receive nominal lump-sum transfers $T = (\mu - 1)M_t$ in market 2 of period $t$. We let $V(m)$ denote the continuation payoff of an agent who has $m > 0$ nominal balances at the start of a period.

Competitive banks exist that accept one-period monetary deposits and make one-period monetary loans at the start of market 1, before the goods market opens—one-period contracts are optimal due to quasi-linear preferences (see BCW, 2007). Banks are profit maximizing firms, owned by households. Consumers can borrow money from banks to supplement their transaction balances, and workers can deposit their idle balances in a bank. Borrowers are charged interest payments on their loans, in market 2; let $r_H$ denote the nominal interest rate on consumer loans. Depositors are compensated with interest payments on their deposits, in market 2; let $r$ denote the nominal interest rate on deposits. Following BCW, we assume that banks can force repayment of loans at no cost, and that banks can engage in intermediation because only they can operate a technology that keeps records of financial histories (but not trading histories). However, unlike BCW, we do not assume that intermediation is a costless process. Instead, we assume that to transform deposits into loans banks must operate a loans' production technology $F$ that uses labor inputs. Hence, in market 1 banks hire workers on a competitive labor market and compensates them in market 2, when loans are returned and any bank's profit is distributed to households as dividends. For simplicity, we will develop the main results under the conjecture that no one makes a deposit unless they are compensated for it, i.e., deposits are made only if $r > 0$. This is equivalent to assuming some (infinitesimal) cost from holding a deposit account. Subsequently, we will relax this assumption and consider the case when deposits are made even if $r = 0$.

# 3 Equilibrium allocations

We focus on a symmetric stationary monetary equilibrium where all agents of the same type follow identical strategies and where real variables are constant over time. Consider a generic date $t$. Let $\phi$ denote the real price of money in market 2. Variables corresponding to the following date $t+1$ are indexed by a prime. In a stationary equilibrium real money balances are time invariant, so we have $\phi M = \phi' M'$, which implies that inflation is pinned down by the rate of money growth because:

$$\frac{\phi}{\phi'} = \frac{M'}{M} = \mu$$

Consider a stationary equilibrium on a generic date $t$ when banks are operating. Let $W(m, l, d)$ denote the continuation payoff of an agent who, at the start of market 2, has $m$ balances, $l$ loans and $d$ deposits. This portfolio depends on the agent's history of trade in market 1. We have:

$$\begin{aligned} W(m, l, d) &= \max_{X,Y,m' \geq 0} U(X) - Y + \beta V(m') \\ \text{s.t. } Y &= X - \phi m - \phi p n_B - \phi(1+r)d + \phi(1+r_H)l + \phi m' - \phi T - \phi\zeta \end{aligned}$$

where $m'$ is money taken into period $t+1$, $T$ are lump-sum transfers and $\zeta$ are banks' dividends. To understand the budget constraint note that only workers can supply labor to a bank and would want to make deposits, while only consumers would want to take out loans. A consumer who borrowed $l > 0$ money in market 1 must repay $(1 + r_H)l$ money in market 2, whereas a worker who deposited $d > 0$ money in market 1 receives $(1 + r)d$ money in market 2. A worker who supplies $n_B$ labor to banks in market 1 gets paid $pn_B$ dollars in market 2 when the wage is $p$. Note that here $p$ corresponds to the price of goods in market 1 because in equilibrium workers must be indifferent between producing goods or working for a bank.

Substituting the constraint into $W$ yields:

$$W(m, l, d) = \phi[m + pn_B + (1+r)d - (1+r_H)l + T + \zeta] + \max_{X,m' \geq 0}[U(X) - X - \phi m' + \beta V(m')]$$

The first order conditions in a monetary equilibrium are:

$$(X) \quad : \quad U'(X) = 1 \;\Rightarrow\; X = X^* \tag{1}$$

$$(m') \quad : \quad \phi = \beta V'(m') \tag{2}$$

The envelope conditions are:

$$\frac{\partial W}{\partial m} = \phi, \qquad \frac{\partial W}{\partial d} = \phi(1+r), \text{ and } \qquad \frac{\partial W}{\partial l} = -\phi(1+r_H)$$

We also have $\dfrac{\partial W}{\partial n_B} = \phi p$.

Moving backwards in the sequence of markets, the expected value of holding $m$ money balances at the start of market 1 when the price is $p$ is:

$$V(m) = \max_{q,l} \; \alpha[u(q) + W(m - pq + l, l, 0)] + \max_{n,n_B,d} \; (1-\alpha)[-c\,(N) + W(m - d + pn, 0, d)]$$

Note that $N = n + n_B$ is the total labor supply of a worker and $W$ is a function of the labor supplied to banks in market 1, $n_B$.

We now consider workers' and consumers' decisions, separately. The worker's problem when banks are operating is:

$$\begin{aligned} \max_{n,n_B,d} \quad & -c(n + n_B) + W(m - d + pn, 0, d) \\ \text{s.t.} \quad & d \le m \text{ and } n + n_B = N \end{aligned}$$

The first-order conditions are:

$$c'(N) = p\phi \tag{3}$$

$$\phi r = \lambda_d \tag{4}$$

where $\lambda_d$ is the multiplier on the deposit constraint. The labor supply $N$ is determined by the real wage $p\phi$, and is independent of the worker's financial decisions. Moreover, (4) implies that if $r > 0$, then the deposit constraint is binding and $d = m$. If $r = 0$, then the agent would be indifferent between keeping the money or depositing in the bank only in the absence of costs for making deposits. We will work under the conjecture that the agent does not make a deposit if $r = 0$ and consider the case in which she does, in a later section.

The problem for a consumer when banks are operating is:

$$\begin{aligned} \max_{q,l} \quad & u(q) + W(m - pq + l, l, 0) \\ \text{s.t.} \quad & pq \leq m + l \end{aligned}$$

The first-order conditions are:

$$\begin{aligned} u'(q) - \phi p - p\lambda = 0 \\ \phi r_H = \lambda \end{aligned}$$

where $\lambda$ is the multiplier on the budget constraint. These two equations jointly imply:

$$u'(q) = \phi p(1 + r_H)$$

Using the information from the worker's problem, we have:

$$u'(q) = c'(N)(1 + r_H) \tag{5}$$

where $N = n + n_B$. The value $q$ is also a function of $n$ and $n_B$ because the goods market clearing condition is:

$$(1 - \alpha)n = \alpha q \tag{6}$$

Using the envelope conditions, the marginal value of money satisfies:

$$V'(m) = \phi \left[ \frac{\alpha u'(q)}{p\phi} + (1 - \alpha)(1 + r) \right]$$

which, given $u'(q) = \phi p(1 + r_H)$, yields:

$$V'(m) = \phi\,[1 + r + \alpha(r_H - r)] \tag{7}$$

We now can move on to discussing the bank's problem. It is assumed that hiring $\ell > 0$ labor allows the bank to generate $F(\ell)$ real loans, which are measured in terms of goods in market 1.[3] It is assumed that $F(0) = 0$, with $F' > 0$, $F'(0) < \infty$ and $F'' \leq 0$. The bank's profit maximization problem is:

$$\begin{aligned} \max_{L,D,\ell} \quad & L r_H - D r - p\ell \\ \text{s.t.} \quad & L = pF(\ell), \text{ and } L \leq D \end{aligned}$$

where $L$ and $D$ are aggregate loans and deposits, respectively. It should be clear that banks will operate only if there is a positive spread between interest rate on loans and deposits, i.e. $r_H > r$. In that case, the second constraint holds with equality. Hence, conjecturing $r_H > r$, we substitute both constraints in the production function, and the bank's maximization problem becomes:

$$\max_{\ell}\; pF(\ell)(r_H - r) - p\ell$$

The FOC is $F'(\ell)(r_H - r) = 1$. Therefore:

$$r_H - r = \frac{1}{F'(\ell)} \tag{8}$$

Depending on $F$, banks may earn positive profits, which are distributed as dividends $\zeta = p[F(\ell)/F'(\ell) - \ell]$. Note that greater efficiency in financial intermediation results in lower spreads. Spreads, in turn, alter the marginal value of money in (7), which increases with the spread, i.e., money is more valuable when financial intermediation services are *less* efficient because loans are more expensive, so consumers must rely more on cash.

[3] An alternative is to measure loans in terms of the market 2 good. It is preferable to consider the loan amount in terms of market-1 goods because banks make loans that are used to purchase goods in market 1, thus the measure proposed is more convenient.

Finally, note also that the spread is non-negative for all levels of banking. The spread cannot be zero because $F'(\ell) < \infty$ for all values of $\ell$.

Now consider market-clearing. The loan-production technology implies:

$$L = \alpha(pq - m) = pF(\ell) \tag{9}$$

The lending constraint is $L = \alpha(pq - m) = D = (1 - \alpha)m$. This implies that per capita money holdings are equal to per capita expenditure in market 1, so that:

$$\alpha pq = m \tag{10}$$

Therefore, we get:

$$F(\ell) = \alpha(1 - \alpha)q \tag{11}$$

where from labor market clearing we have:

$$\ell = n_B(1 - \alpha)$$

Using (11) and the goods market-clearing condition in (6), in equilibrium we have:

$$(1 - \alpha)^2 n = F(n_B(1 - \alpha)) \tag{12}$$

# 4 Results

In this section, we list the main findings for economies where—as in the typical model of this class—the central bank does not pay any interest on commercial banks' reserves. Subsequently, we will also consider the case where these payments can be made. Last, we discuss the quantitative implications of the model.

## 4.1 Zero interest on reserves

This is the standard assumption in this class of models. Start by noting that the nominal interest rate on an illiquid bond is $i = \frac{\mu - \beta}{\beta}$. We use this definition to determine the equilibrium interest rate on bank deposits.

**Result 1** (**Endogenous spread**). *In equilibrium there is an endogenous spread between loan and deposit rates,* $0 \leq r = r^* < i < r_H$ *with:*

$$r^* := \max\left(i - \frac{\alpha}{F'(\ell)}, 0\right) \quad \textit{and} \quad r_H = \frac{i}{\alpha} - \frac{r^*(1-\alpha)}{\alpha}$$

*Hence, if* $\mu \leq \beta + \frac{\beta\alpha}{F'(0)}$*, then* $r = 0$*.*

To derive this result, note that the marginal value of money must satisfy (2) and (7). The basic Euler equation is thus:

$$\frac{\mu}{\beta} = 1 + r + \alpha(r_H - r) \qquad \Rightarrow \qquad i - r = \alpha(r_H - r) \tag{13}$$

To determine the interest rates use equation (8), which identifies the spread, yielding the candidate deposit rate:

$$r = i - \frac{\alpha}{F'(\ell)} \tag{14}$$

Since agents can trade with cash and are free to hold cash outside of banks, the deposit rate must be positive in a banking equilibrium. Hence, we have $r^*$.[4] If $r = 0$, then $\ell = 0$ since banks are inactive and so the aggregate labor effort is $N = \frac{\alpha q}{1-\alpha}$ (no one works in a bank).

From (13) with $r = r^*$ we obtain:

$$r_H = \frac{i}{\alpha} - \frac{r^*(1-\alpha)}{\alpha}$$

[4] This is what happens in BCW at the Friedman rule, where rates on loans and deposits are zero. There, it is irrelevant if banks do or not operate. Here, it is consequential because banks employ workers, which has general-equilibrium effects. For this reason, in the Appendix we also study the case where sellers deposit their money when $r = 0$.

Note that $r_H > i$ so borrowers always pay a premium above the yield $i$ on an illiquid bond. When $r^* > 0$, the loan rate is:

$$r_H = i + \frac{1-\alpha}{F'(\ell)} \tag{15}$$

The interest rates on deposits and on loans are governed by the productivity of the banking sector, $F'$. In particular, one can think of a model with costless banking as a special case of ours, where banks are infinitely productive at any level of activity. In that case, $r^* = i$ and $r_H = r = i$. The link between banks' productivity and interest rates on deposits has the following further important implication.

**Corollary 1** (**Banking and deflations**). *The banking equilibrium breaks down for inflation rates sufficiently close to the Friedman rule.*

When banking absorbs labor resources, banking breaks down for inflation rates sufficiently close to $\beta$. The reason is that $r < i$. Hence, for $i > 0$ we have that $r = r^* = 0$ for $\mu \leq \beta + \frac{\beta\alpha}{F'(0)}$, and sellers no longer make deposits. Thus, in economies with less productive banking sectors a banking equilibrium exists only at higher inflation rates. This result stands in contrast with the analysis in BCW. There, banks are active and improve the allocation for any inflation rate above the Friedman rule, while for $\mu \to \beta$ agents are indifferent between cash and deposits and the allocation with banking is equivalent to the allocation attained in a pure-currency economy.

**Result 2.** *Consider $r^* > 0$. In a banking equilibrium, the market 1 allocation satisfies:*

$$u'(q) = c'(N)\left[1 + \frac{i}{\alpha} - \frac{r^*(1-\alpha)}{\alpha}\right] \tag{16}$$

*Thus, everything else equal, economic efficiency increases as the deposit rate $r^*$ increases.*

Using $r_H$ from Result 1 in (5) we obtain (16). To prove the second part of the result, expression (16) implies that $q$ increases with $r^*$ because $N > \frac{\alpha q}{1-\alpha}$, while $u$ is concave and $c$ is convex. Since $r^*$ increases with banks' productivity, then economic efficiency also increases with banks' productivity. This in turn implies that the efficiency level achieved when banking services require labor inputs is always below the efficiency level attained in BCW, where:

$$u'(q) = c'(N)(1 + i)$$

and the aggregate labor effort is $N = \dfrac{\alpha q}{1-\alpha}$, which is less than labor effort in our model where banks also use labor inputs.

Introducing a more realistic model where intermediation services are costly has also a deep implication for the desirability of banking. A fundamental result with costless banking is that financial intermediation *always* improves the allocation and welfare when $\mu > \beta$ (BCW, Corollary 1). Simply put, away from the Friedman rule a pure currency economy would always deliver an inferior allocation. This is no longer true when intermediation absorbs labor resources. There is a threshold inflation level below which active intermediation lowers macroeconomic efficiency. This threshold level depends on the productivity of the banking sector.

**Result 3** (**Banking may reduce welfare**). *If $r^* > 0$ and inflation is sufficiently low, then the banking equilibrium can be Pareto-inferior relative to the monetary equilibrium without banks.*

This result implies that banks can reduce efficiency even if they pay interests on deposits. Banks, in this case, simply soak up too many labor resources—distorting prices—relative to the benefits provided to depositors. To prove it, one needs to show that the benefit from greater consumption (thanks to banks) is less than the cost of greater work effort. Ex-ante welfare is given by:

$$(1 - \beta)V = \alpha u(q) - (1 - \alpha)c(N) + U(X^*) - X^*$$

Banking activity affects ex-ante welfare because it distorts market-1 consumption and labor, and thus the expected trade surplus $\alpha u(q) - (1 - \alpha)c(N)$.

Conjecturing that $i$ is such that depositing money is profitable ($r^* > 0$), denote $q = q_0$ and $N = N_0$ the endogenous variables in the equilibrium without banks, and $q = q_1$ and $N = N_1$ the endogenous variables in the equilibrium with banks. Using the equilibrium condition in Result 2, we have:

$$u'(q_0) = c'(N_0)\left[1 + \frac{i}{\alpha}\right] \qquad \text{with } N_0 = \frac{\alpha q_0}{1-\alpha} \equiv n_0$$
$$u'(q_1) = c'(N_1)\left[1 + \frac{i}{\alpha} - \frac{r^*(1-\alpha)}{\alpha}\right] \qquad \text{with } N_1 = \frac{\alpha q_1 + \ell}{1-\alpha} \equiv n_1 + n_B$$

The second line reveals that we have a discontinuity in $\ell$ and $q$ as we switch from the monetary equilibrium to the banking equilibrium. This is because in the banking equilibrium $F(\ell) = (1-\alpha)\alpha q_1$ from (11), so $\ell$ is bounded away from zero.

There are two possible cases: $q_1 > q_0$, and $q_1 \leq q_0$. Either way, one can prove that $N_1 > N_0$ and therefore $c'(N_1) \geq c'(N_0)$ since $c'' \geq 0$. If $q_1 > q_0$, then it is immediate that $N_1 > N_0$ since $\ell > 0$ in the banking equilibrium. The second case $q_1 \leq q_0$ is only possible if $c'' > 0$ because $r^* > 0$ and $u'' < 0$; here, we again have $N_1 > N_0$.

Given this finding, we can now compare welfare with and without banks. Welfare is higher *without* banks if:

$$\alpha u(q_0) - (1-\alpha)c(N_0) > \alpha u(q_1) - (1-\alpha)c(N_1)$$

If $q_1 \leq q_0$, then banks clearly reduce welfare since agents do not consume more but work more, since $N_1 > N_0$. If, instead, $q_1 > q_0$, then the curvature of the cost function $c$ matters since $N_1 > N_0$ and therefore it is the marginal cost that matters here—a finding that is similar to the one presented in Chiu et al. (2018). Here, welfare with banks is below welfare without banks if the additional utility (generated by the extra consumption $q_1 - q_0$) is less than the additional cost associated with the extra labor effort $N_1 - N_0$. This is likely when the consumption increment is small, which happens if inflation is sufficiently low, as in that case consumption without banks is already close to the efficient amount $q^*$. Notice also that if the cost is linear, then this channel disappears much as it happens in Chiu et al. (2018).

This result stands in contrast to a fundamental result in BCW (Proposition 2), stating that banks always improve the equilibrium allocation away from the Friedman rule. The intuition is that banks improve welfare because they compensate sellers for their idle balances and not because they slacken the cash constraint of buyers (who are indifferent to borrowing). In this manner, everyone brings more liquidity to the market because money never sits idle, and market 1 production increases. As soon as banks stop compensating depositors, then intermediation becomes irrelevant. In BCW, banks are irrelevant only at the Friedman rule because they have no operating costs so that $0 < r = i = r_H$ for $\mu > \beta$. That is: deposits yield as much as an illiquid bond, and loans command that same rate. Instead, in our model intermediation is costly, which implies an endogenous spread exists between loans and deposit rates, $0 < r < i < r_H$. In this case, deposits yield less than an illiquid bond, while loans command a premium over it. For inflation rates sufficiently small, banks end up compensating depositors too little and at the same time charge too high a premium to borrowers, in which case the allocation can be worse than in an economy without banks.

**An example with linear $F$.** Consider a technology that transforms labor into loans at constant marginal rate, $F(x) = Ax$ for $A > 0$. The equilibrium interest rate is:

$$r^* := \max\left(i - \frac{\alpha}{A}, 0\right), \quad \text{and} \quad r_H = \frac{i}{\alpha} - \frac{r^*(1-\alpha)}{\alpha}$$

Hence, $r > 0$ only for inflation rates satisfying $\mu \geq \mu_L := \beta + \dfrac{\beta\alpha}{A}$.

Equation (12) implies:

$$An_B = (1-\alpha)n$$

The Euler equation (16) becomes:

$$u'\left(\frac{(1-\alpha)n}{\alpha}\right) = c'(n + n(1-\alpha)/A)\left[1 + \frac{i}{\alpha} - \frac{r^*(1-\alpha)}{\alpha}\right]$$

This will give us $n$ as a function of the parameters.

It is immediate that when $\mu \leq \mu_L$ the economy is better off without banks, because in this case $r = 0$. Now suppose $\mu > \mu_L$ so $r^* = i - \frac{\alpha}{A} > 0$. Let $i_L$ denote the interest rate when $\mu = \mu_L$.

To determine when the economy is better off with or without banks we must use specific functional forms. Let $u = \ln(q)$ and $q = n$. Then, with banks:

$$\frac{\alpha}{(1-\alpha)n_1} = c'(n_1 + n_1(1-\alpha)/A)\left[1 + i + \frac{1-\alpha}{A}\right]$$

Note that as $A = \infty$ we get back the model in BCW. Without banks, as in Lagos and Wright (2005), we have:

$$\frac{\alpha}{(1-\alpha)n_2} = 1 + \frac{i}{\alpha}$$

The left-hand side of each equation is $1/q$. The quantity consumed $q$ is greater with banks when the right-hand side of the first equation is less than the right-hand side of the second equation when $n_1 = n_2$, i.e.:

$$c'(n + n(1-\alpha)/A)\left[1 + i + \frac{1-\alpha}{A}\right] \;\; \leq \;\; 1 + i/\alpha$$

Substituting for $n_1 = n_2 = \alpha^2/(1-\alpha)(\alpha + i)$ we obtain:

$$c'\left(\frac{(A+1-\alpha)\alpha^2}{A(1-\alpha)(\alpha+i)}\right)\left[1 + i + \frac{1-\alpha}{A}\right] \;\; \leq \;\; 1 + i/\alpha$$

which can be solved for $i$. Note that the inequality above implies that the larger is the employment in the banking sector ($(1-\alpha)n_B$), the greater is the spread between borrowers' and depositors' rates and therefore the less likely it is that the banking sector is desirable. In other words, low inflation rates are not consistent with large banking sectors. Note also that if $c'' = 0$ then banking is always optimal, since we showed before that $1 + i/\alpha - r^*(1-\alpha)/\alpha < 1 + i/\alpha$. The Euler equation under banking becomes:

$$1 + i + \frac{1-\alpha}{A} - \frac{1}{q} = 0$$

and therefore:

$$\frac{dq}{dA} > 0$$

This implies the more efficient is the bank, the more the agent consumes.

If instead $c'' > 0$, then generally we need an interest rate $i > i_L$ for banking to be optimal. To show it, suppose $\mu = \mu_L$ so $i = i_L$ and $1+i/\alpha - r^*(1-\alpha)/\alpha = 1+i/\alpha$. Hence, banks improve upon the equilibrium only if $c' < 1$. This implies a value $i$ sufficiently large when $c'' > 0$.

**An example with non-linear $F$.** Figure 1 provides a numerical illustration for the case of $F'' > 0$ and $c'' > 0$. This allows us to illustrate that $q$ may indeed drop in banking equilibrium relative to monetary equilibrium, at low inflation rates.

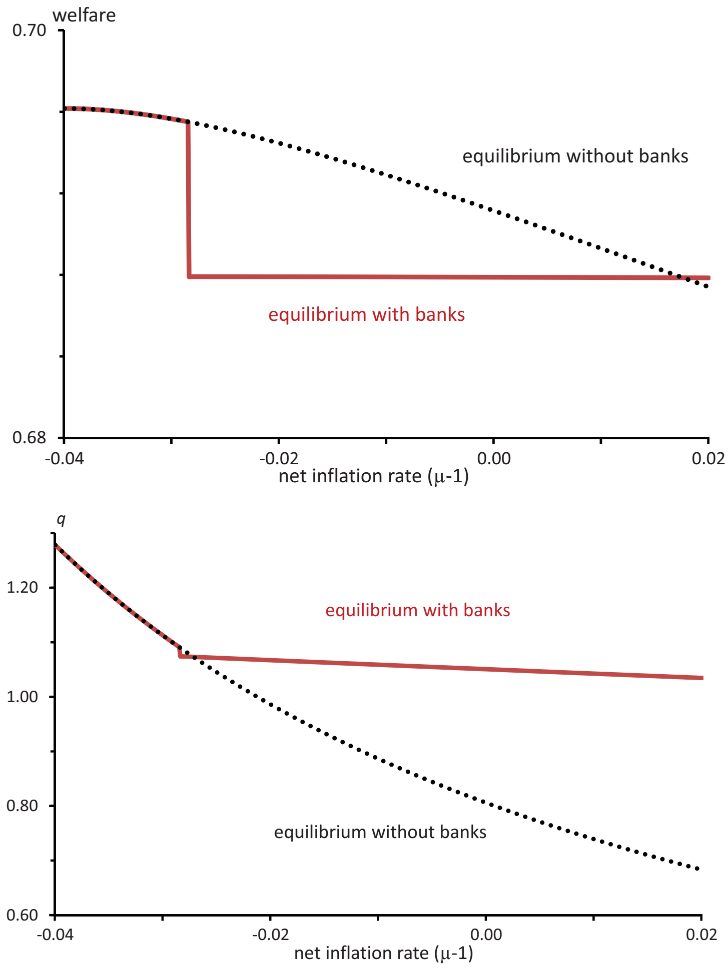


Figure 1: Welfare and market 1 consumption in equilibrium with and without banks.

**Notes:** The horizontal axis reports the net inflation rate $\mu - 1$. The vertical axis reports the welfare level (first panel) and the $q$ amount exchanged in market 1 (second panel). The solid line is our model of costly banking, drawn under the assumption that agents do not make deposits if $r^* = 0$. The dotted line is equilibrium without banks, which effectively corresponds to the model in Lagos and Wright (2005). The figures are drawn for $u(q) = ((q+b)^{1-\delta} - b^{1-\delta})/(1-\delta)$, $U(x) = B\ln(X)$ so that $X^* = B$, $c(N) = N^\psi/\psi$, and a nonlinear loan production function $F(\ell) = A\ln(1+\ell)$. We set $\psi = 1.1$, $B = 2.73$, $\delta = 0.99999$, $\beta = 0.96$, and $b = 0.00001$, which correspond to the values reported in the calibration exercise in the Appendix, while we use smaller values $\alpha = 0.063$, $A = 5.327$ (one-tenth of the values calculated in the Appendix) in order to more prominently display the consumption drop in the figure.

At the Friedman rule ($i = 0$ and deflation at 4% rate) banks cannot offer positive deposit rates and remain inactive. Here, the allocation is efficient. As inflation increases, the nominal interest rate $i$ rises above zero, but initially banks cannot yet offer a positive deposit rate because the prospective cost from hiring labor would cut into profits. Hence, banks remain inactive until we reach an inflation rate of about -3% at which point banks can offering a positive deposit rate. Here, the banking equilibrium and the monetary equilibrium generate different allocations. Initially, welfare drops relative to the monetary equilibrium (top panel). This discrete drop occurs because in this example market-1 consumption drops a bit (bottom panel) and, in addition, agents have additional labor disutility as $\ell$ climbs above zero. The drop in consumption occurs because the increase in labor effort due to banks hiring raises marginal cost of labor ($c$ is convex), hence increasing the price of market-1 consumption. This is similar to the pricing externality through the marginal cost channel in Chiu et al. (2018). However, as $i$ keeps rising, consumption in the monetary equilibrium drops below consumption in the banking equilibrium. This does not immediately erase the welfare disparities as the increase in labor disutility still dominates the extra consumption utility. Eventually, as $i$ grows further this imbalance disappears and welfare in the banking equilibrium overtakes welfare without banks.

**Reducing the size of the banking sector.** Since at low inflation rates banks may end up reducing welfare, instead of increasing it, one could consider policies that aim at reducing the size of the banking sector or completely eliminate it.[5] A first possibility is a tax on banks' hiring. This policy can effectively reduce to zero the inefficiency from banking by making it impossible for banks to raise funds by selling deposits. To see why, let us say for simplicity that the bank must pay a proportional tax $h \in (0, 1)$ in addition to the worker's salary. The bank's maximization problem becomes:

$$\max_{\ell} pF(\ell)(r_H - r) - p(1 + h)\ell$$

The FOC is $F'(\ell)(r_H - r) = 1 + h$. Therefore:

$$r_H - r = \frac{1 + h}{F'(\ell)} \qquad \Rightarrow \qquad r = i - \frac{\alpha(1 + h)}{F'(\ell)}.$$

[5]We thank an anonymous referee for suggesting we discuss the specific policies in this section.

This policy lowers the profitability of banks at the margin, and therefore can lower deposit rates all the way down to zero. Since we need $r > 0$ for agents to make deposits, it follows that a policy of taxing banks' hiring can shut down the banking sector at low rates of inflation.

Another possibility would be to restrict the lending rate. If banks reduce welfare, then restricting the lending rate would reduce banks' profitability, thus making it impossible to offer non-zero returns on deposits. This would also prevent banks from raising funds. Now note that these two policies are likely to backfire if banks are welfare-increasing because, as explained in BCW, the benefit does not come from buyers being able to borrow and consume more, but because sellers can obtain a positive return on their idle balances. It follows that any policy that directly or indirectly lowers the deposit rate is suboptimal if banks are welfare-enhancing. For this reason, we now study the case when banks are paid interest on their reserves and, therefore, can offer higher deposit rates.

## 4.2 Interest on reserves

In this section we study how allocations change when the central bank pays interests on banks' reserves and finances them with lump-sum transfers. Let $\eta$ denote the interest on reserves paid by the central bank, which adjusts $T$ to maintain the inflation rate at a desired level. In line with BCW, ours is a model of narrow "banking" in that banks cannot create more loans than the deposits they have, so the required reserve ratio is 100 percent of deposits. Hence, the bank's profit condition changes as follows:

$$\begin{aligned} \max_{L,D,\ell} \quad & Lr_H - D(r-\eta) - p\ell \\ \text{s.t.} \quad & L = pF(\ell) \text{ and } L \leq D \end{aligned}$$

The second constraint holds with equality if $r_H > r - \eta$, so we can think of interests on reserves as a subsidy on deposits payments. If the second constraint holds with equality and we substitute both constraints in the production function, the bank's maximization problem becomes:

$$\max_{\ell} \; F(\ell)(r_H - r + \eta) - p\ell$$

The FOC for the problem above is:

$$F'(\ell)(r_H - r + \eta) = 1$$

Therefore:

$$r_H - r = \frac{1}{F'(\ell)} - \eta \tag{17}$$

Note that paying interest on reserves reduces the spread.

**Result 4.** *Paying interest on reserves has the following effect: in equilibrium,* $0 \leq r^{**} = r < i < r_H$ *with:*

$$r^{**} := \max\left(i - \frac{\alpha}{F'(\ell)} + \alpha\eta, 0\right) \quad \textit{and} \quad r_H = \frac{i}{\alpha} - \frac{r^{**}(1-\alpha)}{\alpha}$$

*We have* $r^{**} \geq r^{*}$*, and this improves market 1 consumption relative to banking equilibrium without interest on reserves.*

To derive this result, note that the basic Euler equation is still:

$$\frac{\mu}{\beta} = 1 + r + \alpha(r_H - r) \qquad \Rightarrow \qquad i - r = \alpha(r_H - r)$$

so substituting $(r_H - r)$ from equation (17), which identifies the spread when the interest rate $\eta$ is paid on reserves, we obtain the deposit rate:

$$r = i - \frac{\alpha}{F'(\ell)} + \alpha\eta$$

The deposit rate must be non-negative, in equilibrium. Hence we have $r^{**}$. From (13) with $r = r^{**}$ we obtain:

$$r_H = \frac{i}{\alpha} - \frac{r^{**}(1-\alpha)}{\alpha}$$

The Euler equation remains:

$$u'(q) = c'(N)\left[1 + \frac{i}{\alpha} - r^{**}\frac{1-\alpha}{\alpha}\right]$$

Recall that in equilibrium $F(\ell) = \alpha(1-\alpha)q$ and that $N = \frac{\alpha q + \ell}{1-\alpha}$.

We now prove that $r^{**} \geq r^*$. In doing so, for convenience let $q_\eta, N_\eta, \ell_\eta$ denote the equilibrium quantities when there are interests on reserves, and $q, N, \ell$ when there are no interest on reserves. These variables satisfy respectively

$$u'(q) = c'(N)\left[1 + \frac{i}{\alpha} - r^{*}\frac{1-\alpha}{\alpha}\right] \qquad \text{with} \qquad N = \frac{\alpha q + \ell}{1-\alpha}$$

$$u'(q_\eta) = c'(N_\eta)\left[1 + \frac{i}{\alpha} - r^{**}\frac{1-\alpha}{\alpha}\right] \qquad \text{with} \qquad N_\eta = \frac{\alpha q_\eta + \ell_\eta}{1-\alpha}$$

where in equilibrium $F(\ell) = \alpha(1-\alpha)q$ and $F(\ell_\eta) = \alpha(1-\alpha)q_\eta$. We have two cases to consider:

- Case 1: $q_\eta > q$. Since $F' > 0$ and $F(\ell) = \alpha(1-\alpha)q$, we have $\ell_\eta > \ell$. This immediately implies $N_\eta > N$; so $c'(N_\eta) \geq c'(N)$ since $c'' \geq 0$. It follows that since $u'(q_\eta) < u'(q)$ we must have $r^{**} > r^*$.

- Case 2: $q_\eta \leq q$. Here $\ell_\eta \leq \ell$. It follows that $F'(\ell_\eta) \geq F'(\ell)$ since $F'' \geq 0$, which if banks are active directly implies $r^{**} > r^*$ as $\eta > 0$.

Now we prove that $q_\eta > q$. Suppose, by means of contradiction, that $r^{**} > r^*$ but $q_\eta \leq q$. We have two cases. If $c'' = 0$, then it is clear that $q_\eta \leq q$ implies a contradiction since $u'' < 0$. If $c'' > 0$, instead, $q_\eta \leq q$ implies $\ell_\eta \leq \ell$ and therefore $N_\eta \leq N$. In turn, this implies $c'(N_\eta) \leq c'(N)$. Since $u'(q_\eta) \geq u'(q)$ by concavity, we need $r^{**} < r^*$, which contradicts the earlier result. Therefore, even in this case we can only have $q_\eta > q$. This completes the proof.

**Interests on reserves vs. monetary equilibrium.** Result 4 indicates that paying interests on reserves improves market 1 consumption in a banking equilibrium, as compared to paying no interest on reserves. The natural question is: does it improve welfare and, in particular, welfare relative to a monetary equilibrium when we are close to the Friedman rule? As we have seen earlier (Result 3), the banking sector can decrease welfare relative to a monetary equilibrium without banks, for low deflations or low inflation rates. In this section, we have shown that paying interest on reserves can raise the quantity consumed on market 1, but also labor effort and, therefore, labor costs. Therefore it is unclear whether paying interests on reserves raises welfare above monetary equilibrium or not. Intuitively, we expect that it should not reverse the result that banks may lower welfare. The reason is simple: by paying a small interest on reserves a banking equilibrium exists for inflation levels that are even smaller than when interest on reserves are zero. Therefore, welfare can fall below the monetary equilibrium level, which is the welfare level attained when banks have zero interests on reserves.

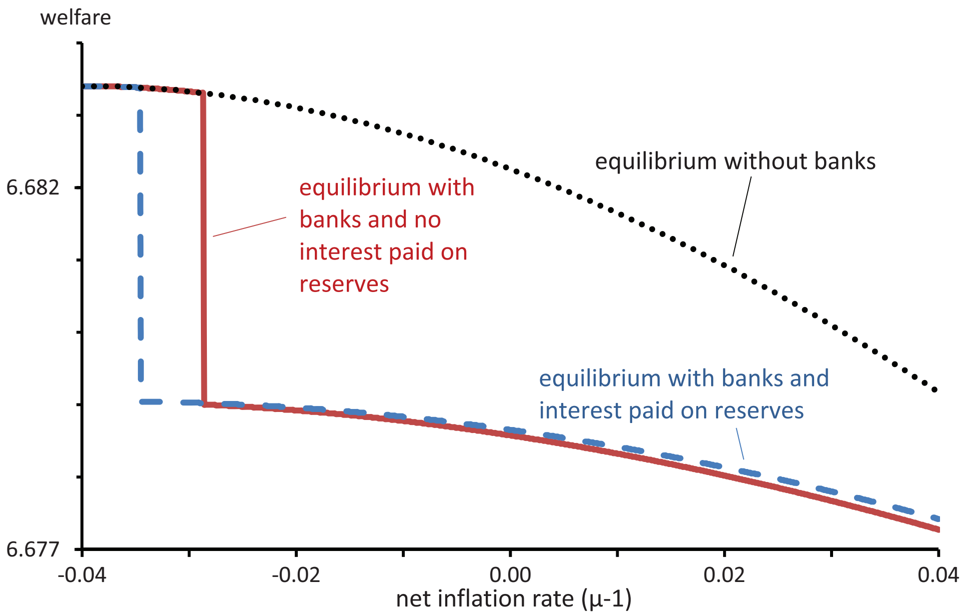


Figure 2: Welfare with and without banks, with and without interests on reserves.

**Notes:** The horizontal axis reports the net inflation rate $\mu - 1$. The vertical axis reports the welfare level. The solid line is our model of costly banking, drawn under the assumption that agents do not make deposits if $r^* = 0$. The dashed line is banking equilibrium when there are $\eta = 0.01$ interests on reserves being paid. The dotted line is equilibrium without banks, which effectively corresponds to the model in Lagos and Wright (2005). The figures are drawn for $u(q) = ((q+b)^{1-\delta} - b^{1-\delta})/(1-\delta)$, $U(x) = B\ln(X)$ so that $X^* = B$, $c(N) = N^{\psi}/\psi$, and a linear loan production function $F(\ell) = A\ell$. We set $\alpha = 0.63$, $A = 53.06$, $\psi = 1.1$, $B = 2.74$, $\delta = 0.99999$, $\beta = 0.96$, and $b = 0.00001$, which are the values reported in the calibration exercise conducted in Section 5.

To illustrate this issue consider Fig. 2, which illustrates welfare under three scenarios (parameters and functional forms in the notes to the figure). The solid line is our model of costly banking, drawn under the assumption that agents do not make deposits if

$r^* = 0$. The dashed line is the banking equilibrium when there are interests on reserves being paid, with $\eta = 0.01$. The dotted line is the equilibrium without banks (monetary equilibrium). For small deflations and low inflation rates in this figure, banks cannot pay a positive return on deposits so the three lines coincide. As inflation increases, when interests are paid on reserves banks become active sooner as compared to the case when no interests on reserves are paid. Paying interest on reserves raises the deposit rate above zero at low inflation rates, thus spurring agents to deposit money into banks. The monetary equilibrium is Pareto-superior to the banking equilibrium for the inflation rates considered in the illustration. This suggests that paying interest on reserves might actually lower welfare close to the Friedman rule because it induces bank activity that would not occur otherwise. As inflation increases further, then banks become active even if no interest on reserves are paid. Here, welfare is larger when interest on reserves is paid as compared to not paid, but still below the welfare without banks until inflation rises sufficiently (not seen in the figure).

# 5 Quantitative analysis

To calibrate the relevant parameters, we focus on a yearly model of the United States for the sample period 1965–2010. All data except for money supply are from the St. Louis Fed *FRED Database*. Interest rates are annualized. The nominal interest rate $i$ is the yield on 3-month treasury bills, the nominal price level $P$ is CPI for all items, aggregate nominal output $PY$ is nominal GDP, the deposit rate $r$ is the yield on 3-month certificates of deposits and the lending rate $r_H$ is the bank prime loan rate.[6] The nominal money supply $M$ is sweep-adjusted M1.[7]

To facilitate comparison with other studies based on Lagos and Wright (2005), we adopt the following functional forms: $u(q) = ((q+b)^{1-\delta} - b^{1-\delta})/(1-\delta)$ with $\delta = 0.99999$ and $b = 0.00001$ so we have approximately unit-elastic preferences in both markets; $U(x) = B\ln(X)$ so that $X^* = B$; and $c(N) = N^\psi/\psi$ with $\psi = 1.1$.[8] For the bank's

[6] Specifically, we use the following *FRED* series: TB3MS for 3-month t bills; CPALTT01USQ661S for CPI; GDP for nominal GDP; IR3TCD01USQ156N for CD rates; MPRIME for the prime rate.

[7] We use the series M1S from Cynamon, Dutkowsky and Jones, "Sweep-Adjusted Monetary Aggregates for the United States." Online document available at http://www.sweepmeasures.com, December 27, 2007. The methods used to create the sweep-adjusted data are described in Cynamon et al. (2006).

[8] Studies based on Lagos and Wright (2005) usually assume linear disutility. Setting $\psi = 1$ would have

problem, we consider a constant return to scale production function $F(\ell) = A\ell$ (we discuss a robustness check for a DRS production function in the Appendix). The assumption of constant returns to scale is convenient for two reasons. First, it matches the assumption on the good's production technology, which is also CRS. Second, it is consistent with the notion that we work with a representative bank in competitive markets, since the bank earns zero profits in equilibrium.

We set $\beta = 0.96$, and for the remaining parameters $\alpha$, $A$ and $B$, we simultaneously match the theoretical expressions for money velocity, interest elasticity of money demand and interest rate spread with their empirical counterparts for the years 1965-2010. Derivations are in the Appendix. The calibration yields the following results: $A = 53.06$, $\alpha = 0.63$ and $B = 2.74$. In what follows, we use these parameter values to quantify the welfare implications of costly banking as well as for the welfare cost of inflation.

One may ask if our numerical exercise matches other features of the U.S. macroeconomy and, in particular, the banking sector employment share and its contribution to GDP. In the model, we find the share of labor employed in the banking sector, relative to total labor, is about 0.7% whereas the value added of banking is 7.2%. In the data, the corresponding statistics are measured at 1.9% and 7% respectively.[9]

## 5.1 Do banks improve social welfare?

The payoff function $V$ can be written as:

$$V(m) = \alpha\left[u(q) + W(m - pq + l, l)\right] + (1 - \alpha)\left[-c\left(N\right) + W(m - d + pn, d)\right]$$

Fixing $\mu$, let $q_\mu$ and $N_\mu$ denote equilibrium quantities; use (1), (6), $T = m' - m$ and the expression for $\zeta$ to define ex-ante equilibrium welfare by $V_\mu$, where:

$$(1 - \beta)V_\mu = \alpha u(q_\mu) - (1 - \alpha)c(N_\mu) + U(X^*) - X^* \tag{18}$$

effectively no impact on the calibration results but it would prevent the study of cases where Result 3 does not hold, i.e., where banks do not necessarily improve welfare.

[9]The share of labor employed in the banking sector is measured using data for employment in credit intermediation and related activities/total employment, both from the *Bureau of Labor Statistics*, 1990–2010. Value added of finance is instead measured using data from *FRED* Finance and Insurance as a percentage of GDP for the period 2005–2010.

Inflation $\mu$ affects ex-ante welfare because it distorts market-one consumption and labor, and thus the expected trade surplus $\alpha u(q) - (1-\alpha)c(N)$.

Figures 3–4 illustrate the effect of banking on social welfare, when banking activities are costly and not. The graphs plot the difference in ex-ante welfare levels with and without active banks as a function of the net inflation rate $\mu - 1$ in an economy. We work under the assumption that agents do not make deposits wen $r^* = 0$. The solid line reports data for the model with costly financial intermediation, whereas the dashed line reports data for the model in BCW. Figure 3 also reports the equilibrium deposit rate (right vertical axis). There are two important observations regarding banks, when their activity absorbs labor resources.

**Observation 1:** *If the inflation rate is sufficiently small, then banks reduce welfare.*

Consider the solid line in Figure 3. If $\mu$ is slightly above the Friedman rule, then banks offer zero deposit rates. In this scenario, agents simply hold cash instead of depositing money in the bank. Here, welfare with or without banks is identical so the solid line coincides with the x axis. As inflation increases, banks become active because they offer a deposit rate that is small, but positive. This generates a drop in welfare because banks do provide small benefits to depositors—which is the main reason for the welfare-enhancing effects of banking in this class of models—but *also* soak labor resources away from productive activities. Overall, this harms macroeconomic efficiency. As inflation increases, banks improve welfare compared to a pure-currency economy—at 7% inflation for the calibrated model.

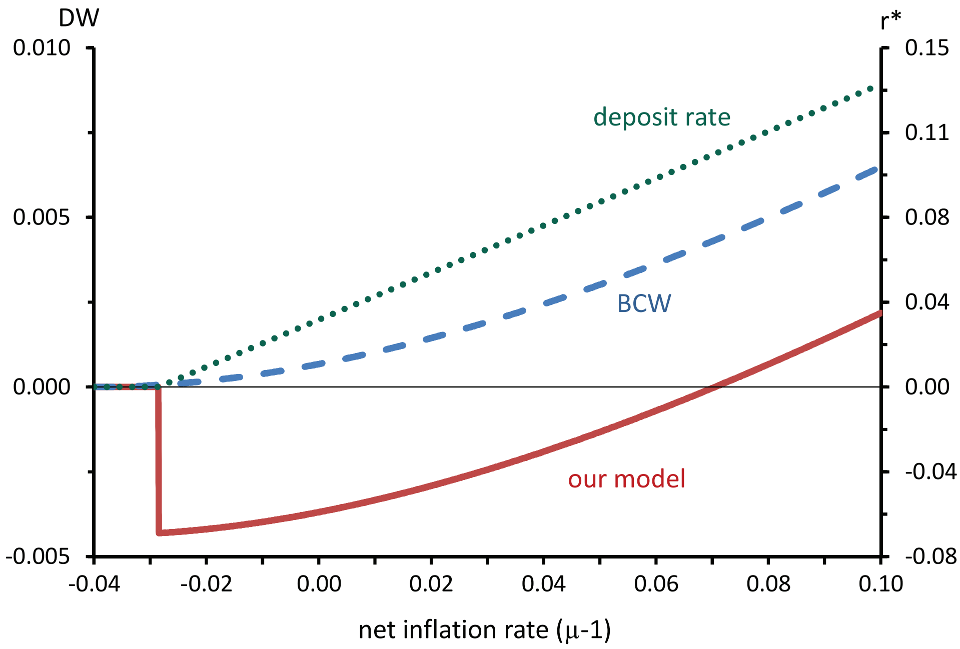


Figure 3: How banks affect welfare at low inflation

**Notes:** The horizontal axis reports the net inflation rate $\mu - 1$. The left vertical axis reports the measure DW, which is the difference in ex-ante welfare levels with and without banks, for the two cases of our model of costly banking (solid line, drawn under the assumption that agents do not make deposits if $r^* = 0$) and the costless banking model in BCW (dashed line). The right vertical axis reports the equilibrium net deposit rate $r^*$ in our model (dotted line).

By means of comparison, Figure 3 also reports welfare differences (in percent) in the calibrated model where banking activity uses no resources at all (as in BCW, that is). In that case, deposit rates are *always* strictly positive (not reported in the figure) so banking activities can only raise welfare, compared to an economy where banks are inactive. The dashed line is therefore monotonically increasing.

The lesson is that incorporating the realistic feature that intermediation is a costly activity reverses the result that banks are welfare improving for all levels of inflation above the Friedman rule.

**Observation 2:** *If the inflation rate is sufficiently high, then banks increase welfare.*

Consider the solid line in Figure 4. As we move away from low inflation scenarios, the welfare gain associated with banking activity grows as depositing money in the bank is a way to partially deflect the inflation tax. The magnitude of this effect depends on the bank's efficiency parameter.

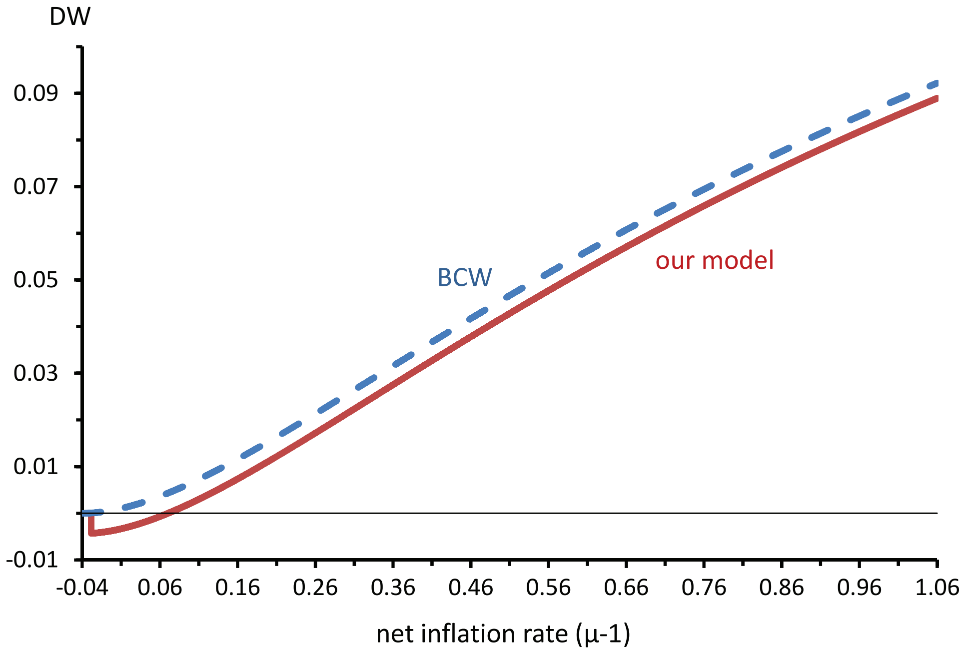


Figure 4: How banks affect welfare at high inflation

**Notes:** The horizontal axis reports the net inflation rate $\mu - 1$. The left vertical axis reports the measure DW, which is the difference in ex-ante welfare levels with and without banks, for the two cases of our model of costly banking (solid line, drawn under the assumption that agents do not make deposits if $r^* = 0$) and the costless banking model in BCW (dashed line).

## 5.2 The welfare cost of inflation

We use a standard measure for the welfare cost of inflation for a representative agent, which is the compensating variation approach proposed in Lucas (2000). The welfare cost of inflation is the percentage adjustment in consumption (in both markets) that leaves the representative agent indifferent between (fully anticipated) inflation $\mu > \beta$ and a lower rate $z \geq \beta$. We use (18) to define ex-ante adjusted welfare $\bar{V}_z$ as:

$$(1-\beta)\bar{V}_z = \alpha u(\bar{\Delta}_z q_z) - (1-\alpha)c(N_z) + U(\bar{\Delta}_z X^*) - X^* \quad (19)$$

The welfare cost of $\mu$ instead of $z$ inflation is the value $\Delta_z = 1 - \bar{\Delta}_z$ that satisfies $V_\mu = \bar{V}_z$.

Table 1 and Figure 5 illustrate the effect of banking on the welfare cost of inflation, when banking activities are costly and not.

**Observation 3:** *Accounting for banking costs raises the welfare cost of small inflation significantly.*

Table 1 reports the welfare cost of inflation as opposed to the Friedman rule, comparing it with the results for (i) costless banking as in BCW and (i) no banking as in Lagos and Wright (2005) using the same calibrated parameters. It also provides the results of a sensitivity analysis in which we use alternative values for the parameter $\psi$.

| | $\psi = 1.1$ | | | $\psi = 1.5$ | | | $\psi = 2$ | | |
|---|---|---|---|---|---|---|---|---|---|
| **Inflation** | 5% | 10% | 20% | 5% | 10% | 20% | 5% | 10% | 20% |
| Costly banking (our model) | 0.20 | 0.28 | 0.52 | 0.18 | 0.24 | 0.41 | 0.17 | 0.22 | 0.34 |
| Costless banking (BCW) | 0.07 | 0.15 | 0.40 | 0.05 | 0.11 | 0.29 | 0.04 | 0.08 | 0.22 |
| No banking (LW) | 0.16 | 0.34 | 0.85 | 0.11 | 0.25 | 0.62 | 0.09 | 0.19 | 0.47 |

Table 1: Welfare costs of inflation relative to the Friedman rule

Figure 5 plots the welfare cost of inflation (in percent) for a wider inflation range.

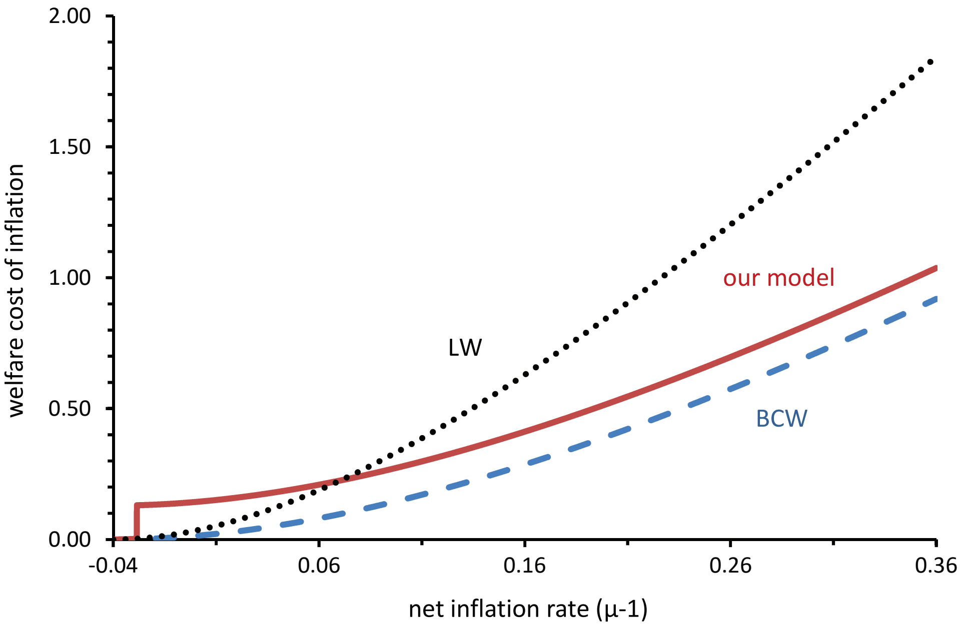


Figure 5: Welfare cost of inflation relative to the Friedman rule

**Notes:** The horizontal axis reports the net inflation rate $\mu - 1$. The vertical axis reports the percentage welfare cost of inflation, relative to the Friedman rule. The solid line is our model of costly banking (drawn under the assumption that agents do not make deposits if $r^* = 0$), the dashed line is the costless banking model in BCW, and the dotted line is the model without banks in Lagos and Wright (2005).

Table 1 and Figure 5 demonstrate that accounting for banking costs has significant quantitative ramifications for the welfare cost of inflation. Although the cost of 10% inflation remains in the order of a fraction of one percent of consumption, as in most models of representative agents, the welfare cost of 5% inflation increases almost three-fold compared to the model where banking is costless, and almost doubles at 10% inflation.

That is because in the model where banks are costless, inflation distorts *only* saving decisions—hence consumption decisions. Instead, when the model accounts for banking costs, inflation also affects the return that depositors receive from their funds—lower than in the model where banks are costless. These two negative effects combine to push up the inefficiency associated with inflation, as compared to the model with costless banking. Such effects decrease at higher inflation rates, because real savings are low in both models and the return to depositors is similar. Note also that at 5% inflation the welfare cost of inflation is higher in an economy with costly banking as compared to a cash economy. This happens because in our calibrated model banks do not increase welfare at that inflation level. This result is reversed at higher inflation rates. Table 1 also shows that higher values for the disutility parameter $\psi$ do not change the results in a significant way. They do however lower the welfare cost of inflation slightly. As explained in the Appendix, the results in Table 1 hold also when agents deposit in the bank even if $r = 0$.

Finally, we note that all of the previous results still hold when we consider a decreasing returns to scale technology for banks. In a separate calibration where $F(\ell) = A\ln(1+\ell)$ we obtain $A = 53.27$, $\alpha = 0.63$ and $B = 2.73$ and figures qualitatively identical to Figures 3-6; the welfare costs of inflation are similar to the ones obtained with a CRS technology. A table with the welfare cost of inflation values for the DRS case is in the Appendix.

To check the robustness of the model, we conducted an additional calibration exercise in which we calibrate $\psi$ instead of just conducting a sensitivity analysis, as we did in Table 1.[10] The calibration is conducted using the DRS technology $F(\ell) = A\ln(1+\ell)$ to match the theoretical share of labor employed in the banking sector in the model $n/N$ with the empirical share of labor in the credit sector, as measured by the *Bureau of Labor Statistics* data. Data are available for the period 1990-2019; the maximum and minimum shares of employment in the credit sector in that sample are 2.04% and 1.68% respectively.[11] In this case, we find $\psi = 1.001, \alpha = 0.53, A = 4.04, B = 4.64$ and $\psi = 1.001, \alpha = 0.53, A = 4.04, B = 4.48$ for employment shares of 1.68% and 2.04% respectively.

[10] We thank an anonymous referee for suggesting this additional analysis. In order to calibrate $\psi$ we must use a nonlinear production function. This implies that, first, we must constrain $\psi$ to be larger than one, and, second, that the results in this calibration exercise are not directly comparable to the ones reported in the previous Table 1, which are obtained using a CRS production function.

[11] The *Bureau of Labor Statistics* also reports data on employment in the entire financial sector, a category that we think would be too broad as compared to the role of banks in our model.

Table 2 reports the findings in terms of the welfare cost of inflation, showing that the welfare cost is slightly higher as employment in the banking sector increases slightly. The reason for this is that as the size of the banking sector increases banking costs also increase, which magnifies the negative effects that inflation has on consumption, through the general equilibrium effects on prices.

| | **Share of labor hired by banks** | | | | | |
|---|---|---|---|---|---|---|
| | 1.68% | | | 2.04% | | |
| **Inflation** | 5% | 10% | 20% | 5% | 10% | 20% |
| Costly banking (our model) | 0.13 | 1.22 | 1.34 | 0.13 | 1.26 | 1.38 |
| Costless banking (BCW) | 0.04 | 0.09 | 0.24 | 0.04 | 0.09 | 0.24 |
| No banking (LW) | 0.13 | 0.28 | 0.67 | 0.13 | 0.29 | 0.70 |

Table 2: Welfare costs of inflation relative to the Friedman rule

# 6 Final comments

Over the last ten years, U.S. households have experienced very low inflation as well as zero yields on demand deposits. In turn, banks have experienced a decrease in deposits while still facing highly negative cash flows due to branch leases and employee compensation. We have thus extended the study of a banking equilibrium in the microfounded model of Berentsen, Camera and Waller (2007) to show why yields on deposits may hit a zero lower bound as inflation shrinks to zero, and to assess the welfare consequences of such policy when the banking sector faces operational costs.

In our model, banks employ labor resources to run their operations, which they hire on a competitive market where they compete with firms. In equilibrium, this gives rise to a spread between interest rates on loans and on deposits, which naturally reflects both the efficiency of financial intermediation and the underlying monetary policy. The main result is that a sufficiently low inflation rate will induce banks to push down to zero the yield on demand deposits, thus removing the primary beneficial effect of banking in this literature. Hence, banks do not necessarily increase welfare under all possible monetary policies. However, when we calibrate the model to U.S. data, we find that these welfare costs are small in terms of how much households would give up to avoid very low inflation (and the need to resort to banks) in favor of the Friedman Rule where banks have no role to play. A natural question is whether or not this result would still hold if banks had an explicit role in intermediating also capital, for example as it happens in Bencivenga and Camera (2011). We leave this question for future research.

# References


Antunes, A., Cavalcanti, T., and A. Villamil (2013). Costly Intermediation and Consumption Smoothing. *Economic Inquiry* 51(1), 459–472.

Aiyagari, S.R., Braun, A.R., and Z. Eckstein (1998). Transaction Services, Inflation, and Welfare. *Journal of Political Economy* 106(6), 1274–1301.

Bencivenga, V., and G. Camera (2011). Banking in a Matching Model of Money and Capital. *Journal of Money, Credit, and Banking* 43(7), 449–476.

Berentsen, A., Camera, G., and C. Waller (2007). Money, Credit and Banking. *Journal of Economic Theory* 135(1), 171–195.

Calomiris, C.W., and D. Nissim (2014). Crisis-Related Shifts in the Market Valuation of Banking Activities. *Journal of Financial Intermediation* 23, 400–435.

Chiu, J., and C.A. Meh (2011). Financial Intermediation, Liquidity, and Inflation. *Macroeconomic Dynamics* 15(1), 83–118.

Chiu, J., Dong, M., and E. Shao (2018). On the Welfare Effects of Credit Arrangements. *International Economic Review* 59(3), 1621–1651.

Cynamon, B.Z., Dutkowsky, D.H., and B.E. Jones (2006). Redefining the Monetary Aggregates: A Clean Sweep. *Eastern Economic Journal* 32(4), 661–673.

Goldfeld, S., and D.E. Sichel (1990). The Demand for Money. In: Friedman, B.M., Hahn, F.H. (Eds.), *Handbook of Monetary Economics*. Elsevier, Amsterdam, 299–356.

Lagos, R., and R. Wright (2005). A Unified Framework for Monetary Theory and Policy Analysis. *Journal of Political Economy* 113(3), 463–484.

Lucas, R., (2000). Inflation and Welfare. *Econometrica* 68(2), 247–274.

Philippon, T. (2015). Has the US Finance Industry Become Less Efficient? On the Theory and Measurement of Financial Intermediation. *American Economic Review* 105(4), 1408–1438.

Philippon, T., and A. Reshef (2013). An International Look at the Growth of Modern Finance. *Journal of Economic Perspectives* 27(2), 73–96.

Schmiedel, H., Kostova, G., and W. Ruttenberg (2012). The Social and Private Costs of Retail Payment Instruments: A European Perspective. European Central Bank Occasional Paper Series # 137.

# Appendix

# A Derivations for CRS technology

Here we consider $F(\ell) = A\ell$ for $A > 0$. This, together with (12), enables us to find an expression for $n_B$ in terms of $n$, which is:

$$n_B = (1-\alpha)n/A \tag{20}$$

Moreover, the Euler equation in (16) for $r > 0$ can be written as:

$$u'(q) = c'(N)\left[1 + i + \frac{1-\alpha}{F'(\ell)}\right]$$

Using the functional forms we chose for $u(q)$, $c(N)$ and $F(\ell)$, the Euler equation becomes:

$$(q+b)^{-\delta} = (n+n_B)^{\psi-1}\left[1 + i + \frac{1-\alpha}{A}\right]$$

This, given (20) and $\alpha q = (1-\alpha)n$ from goods market clearing, gives us an expression for $n$ only as a function of the parameters:

$$\left[\frac{(1-\alpha)n}{\alpha} + b\right]^{-\delta} = \left[\frac{(A+1-\alpha)n}{A}\right]^{\psi-1}\left[1 + i + \frac{1-\alpha}{A}\right] \tag{21}$$

Now we need to pin down $\alpha$, $A$ and $B$ by matching interest elasticity of money demand, money velocity and the interest rate spread with their empirical counterparts. The general formula for interest elasticity of money demand is:

$$\varepsilon_m = \frac{\partial(m\phi)}{\partial i} \times \frac{i}{m\phi}$$

We know $m = pq$ which, together with (3), implies $m\phi = c'(N)q$. Differentiating we get:

$$\varepsilon_m = \left[c'(N)\frac{dq}{di} + qc''(N)\frac{dN}{di}\right] \times \frac{i}{c'(N)q} = \left[c'(N)\frac{dq}{di} + qc''(N)\frac{dN}{di}\right] \times \frac{i}{c'(N)q} \tag{22}$$

From (2) and (7) the Euler equation can be written as $i = \alpha\left[u'(q)/c'(N) - 1\right] + (1-\alpha)r$. Using the implicit function theorem, we have:

$$\frac{dq}{di} = -\frac{-1}{\alpha\frac{u''(q)}{c'(N)}} = \frac{c'(N)}{\alpha u''(q)}$$

$$\frac{dN}{di} = -\frac{-1}{-\alpha\frac{u'(q)c''(N)}{c'(N)^2}} = -\frac{c'(N)^2}{\alpha u'(q)c''(N)}$$

Therefore, the expression for the elasticity of money demand can be written as:

$$\varepsilon_m = \left[\frac{c'(N)^2}{\alpha u''(q)} - \frac{qc'(N)^2}{\alpha u'(q)}\right] \times \frac{i}{qc'(N)} = \frac{ic'(N)}{\alpha q}\left[\frac{1}{u''(q)} - \frac{q}{u'(q)}\right] \tag{23}$$

Given the functional forms for $u(q)$ and $c(N)$, the expression for $\varepsilon_m$ becomes:

$$\varepsilon_m = -\frac{in^{\psi-2}\left(\frac{A+1-\alpha}{A}\right)^{\psi-1}}{\delta(1-\alpha)}\left[b+\frac{(1-\alpha)n}{\alpha}\right]^{\delta}\left[b+\frac{(1+\delta)(1-\alpha)n}{\alpha}\right] \tag{24}$$

where $n$ is defined in (21). The empirical counterpart is estimated using the approach outlined in Goldfeld and Sichel (1990)[12] for which we obtain -0.1123.

Money velocity is $v = PY/M$ where $PY$ is nominal GDP and $M$ denotes nominal money holdings. In the model, nominal GDP has three components: market 1 goods output $p\alpha q$, market 1 banking output $pF(\ell)$ and market 2 nominal output $p_2X$. Therefore, the theoretical expression for money velocity can be derived as follows:

$$v = \frac{PY}{M} = \frac{p\phi\alpha q + p\phi F(\ell) + X}{m\phi}$$

We know the following is true in equilibrium: $U'(X) = 1/B$ from (1), which implies $X = B$ given the functional form $U(X) = \ln(X)$; $p\phi = c'(n+n_B)$ from (3); $\alpha q = (1-\alpha)n$ from (6); $m = \alpha pq$ from (10); $n_B = (1-\alpha)n/A$ from (20). Hence, velocity in equilibrium becomes:

$$v = 2 - \alpha + \frac{B}{n^{\psi}(1-\alpha)\left[\frac{A+1-\alpha}{A}\right]^{\psi-1}} \tag{25}$$

For the empirical counterpart of $v$, we find $v = 5.90$.

Last, the parameter $A$ is chosen to fit the interest rate spread which, given (8), is:

$$r_H - r = 1/A \tag{26}$$

The empirical counterpart for $r_H - r$ is equal to 1.88%, which pins down a value $A = 53.06$. Given the values we set for $\beta$, $\delta$, $b$ and $\psi$, we find $\alpha = 0.63$ and $B = 2.74$ and we can find all the endogenous variables:

$n_B = (1-\alpha)n/A$ from (20);

$N = n + n_B \Rightarrow N = (A+1-\alpha)n/A$;

$\ell = (1-\alpha)n_B$ from labor market clearing;

$q = (1-\alpha)n/\alpha$ from (6);

$X = B$;

$L = (1-\alpha)l$;

$D = \alpha d$;

$p\phi = c'(N) \Rightarrow p\phi = N^{\psi-1}$.

[12] The log of real money balances on each date $t$ $(M_t/P_t)$ is regressed on the date $t$ log of real GDP, nominal interest rates, and one-period lagged real money balances: $\ln m_t = \gamma_0 + \gamma_1 \ln y_t + \gamma_2 \ln i_t + \gamma_3 \ln m_{t-1} + v_t$. To account for first-order autocorrelation in the residuals $v_t$ the Cochrane-Orcutt procedure is used.

# B Robustness check: DRS technology

Here we calculate the welfare cost of inflation when the technology has decreasing returns to scale. We consider $F(\ell) = A\ln(1+\ell)$. This, together with (12), allows us to find an expression for $n_B$ in terms of $n$, which is:

$$n_B = \frac{e^{(1-\alpha)^2 n/A} - 1}{1-\alpha} \tag{27}$$

Using the functional forms we chose for $u(q)$, $c(N)$ and $F(\ell)$, the Euler equation in (16) for $r > 0$ becomes:

$$\left[\frac{(1-\alpha)n}{\alpha} + b\right]^{-\delta} = (n+n_B)^{\psi-1}\left[1 + i + \frac{(1-\alpha)(1+(1-\alpha)n_B)}{A}\right] \tag{28}$$

Given (27) and $\alpha q = (1-\alpha)n$ from goods market clearing, (28) implicitly defines $n$ as a function of only the parameters.
With DRS returns to scale the expression for elasticity in (23) becomes:

$$\varepsilon_m = -\frac{i(n+n_B)^{\psi-1}}{\delta(1-\alpha)n}\left[b + \frac{(1-\alpha)n}{\alpha}\right]^{\delta}\left[b + \frac{(1+\delta)(1-\alpha)n}{\alpha}\right] \tag{29}$$

where $n$ and $n_B$ are defined in (28) and (27). As before, the empirical counterpart is -0.1123.
Given the functional form for $F(\ell)$, the expression for velocity in equilibrium becomes:

$$v = 1 + \frac{A\ln\left[1+(1-\alpha)n_B\right]}{(1-\alpha)n} + \frac{B}{(n+n_B)^{\psi-1}(1-\alpha)n} \tag{30}$$

Last, given (8) the interest-rate-spread is:

$$r_H - r = \frac{1+(1-\alpha)n_B}{A} \tag{31}$$

Given the values we set for $\beta$, $\delta$, $b$ and $\psi$, once we match (23), (30) and (31) with their empirical counterparts we find $\alpha = 0.63$, $B = 2.73$ and $A = 53.27$ and we can find all the endogenous variables of the model. Specifically:

$n_B = (e^{(1-\alpha)^2 n/A} - 1)/(1-\alpha)$ from (27);

$N = n + n_B \Rightarrow N = n + (e^{(1-\alpha)^2 n/A} - 1)/(1-\alpha)$ ;

$\ell = (1-\alpha)n_B$ from labor market clearing;

$q = (1-\alpha)n/\alpha$ from (6);

$X = B$;

$L = (1-\alpha)l$;

$D = \alpha d$;

$p\phi = c'(N) \Rightarrow p\phi = N^{\psi-1}$

$\zeta = p[F(\ell)/F'(\ell) - \ell]$.

Table 3 reports the welfare cost of several inflation rates as opposed to the Friedman rule in this case.

| Inflation | 5% | 10% | 20% |
|---|---|---|---|
| Welfare Cost | 0.19 | 0.28 | 0. 52 |

Table 3: Welfare costs of inflation relative to the Friedman rule (DRS technology)

# C Households make deposits when $r = 0$

In the previous analysis we have assumed that no deposits are made unless depositors are compensated. However, in that case depositors are in fact indifferent between holding on to their money or leaving it in the bank. Here we suppose instead that they deposit all their money into banks, even if $r = 0$.

Under this scenario the model does not change very much. We still have:

$$u'(q) = c'(N)\left(1 + \frac{i}{\alpha}\right)$$

although one has to be careful in comparing it with the no-banks expression because—though this expression resembles the one derived for the model without banks—the allocations are not identical in the two cases. This is due to a market-clearing condition. The labor supply is $N > \frac{\alpha q}{1-\alpha}$ when banks are active, while it would have an $=$ sign with inactive banks. The main implication is that now banks induce general-equilibrium effects even if deposit rates are zero.

To understand why this happens, take two separate steps. Consider first the economy where banks have no operating costs, as in BCW. There, banks influence allocations *only* by providing benefits to depositors, i.e., offering a return on their idle balances. In that economy, allocations with or without banks are identical when deposit rates are zero—which occurs only at the Friedman rule. Now, let us account for banking costs. Here, banks' activity soaks up labor resources and there is an additional effect—through labor markets—even if depositors receive no compensation for their deposits. All other expressions remain the same as the ones developed earlier.

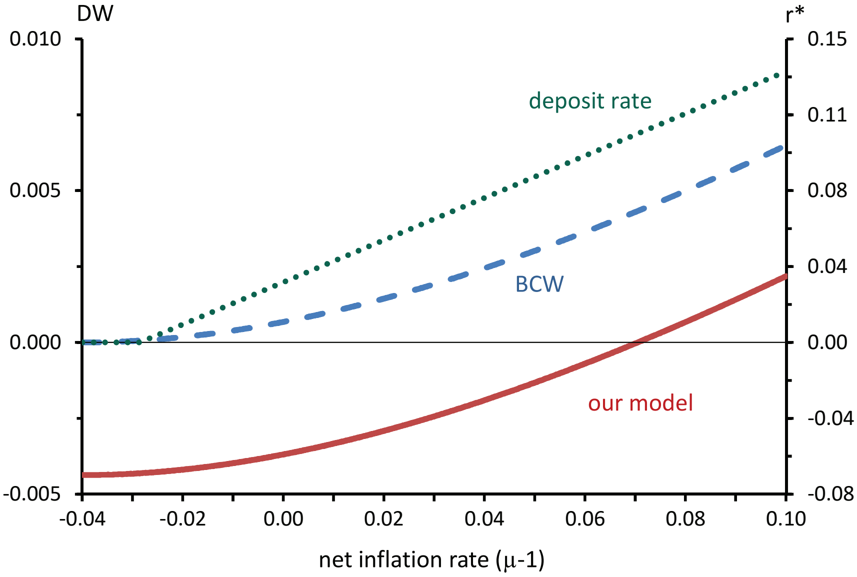


Figure 6: How banks affect welfare when agents always make deposits

**Notes:** The horizontal axis reports the net inflation rate $\mu - 1$. The left vertical axis reports the measure DW, which is the difference in ex-ante welfare levels with and without banks. The solid line is our model (drawn under the assumption that agents make deposits even if $r^* = 0$) and the dashed line is the costless banking model in BCW. The dotted line is the equilibrium deposit rate in our model (scale on right).

The impact of this variation in assumptions is seen at those deflation/inflation levels where $r = 0$. Banking now will induce a welfare loss all the way to the Friedman rule because banks hire workers away from production but do not compensate depositors. This is seen in Figure 6. It should be also clear that in this case there is no effect on the welfare cost of inflation, at positive inflation levels, since zero deposit rates in the calibrated model only occur for deflations close to the Friedman rule.