Optimal Monetary Policy with Interest on Reserves and Capital Over-accumulation

Tai-Wei Hu*

January 2, 2020

Abstract

We propose a model of private and public liquidity provision to study optimal monetary policy implemented with interest on excess reserves (IOER). Banks issue deposits backed by productive capital, and households demand deposits and currency as means-of-payments for decentralized trades in formal and informal markets. Limited commitment of banks implies a pledgeability constraint–deposit issuance constrained by capital and reserve holdings. When financial friction is severe or when productivity is low, banks hold excess reserves in equilibrium and deposit contracts exhibit zero nominal interest rate, and aggregate capital is above its efficient level. When productivity is near the ZLB region, a positive IEOR financed by nominal asset creation is optimal, and, under a fixed inflation target, optimal nominal IOER is increasing in productivity. To implement the constrained efficient allocation, it is optimal to have both a fixed and a proportional liquidity requirement in addition to IOER when productivity is not too high.

JEL Classification: E42, E52, E58

Keywords: Monetary policy, interest on reserves, zero-lower-bound, banking regulations.

1 Introduction

Since the 2008 Global Financial Crisis, there have been at least two significant changes regarding the financial environment relevant for monetary-policy making. First, the nominal interest rate dropped to the zero lower bound (ZLB) and stayed at very low levels for

*University of Bristol, email: taiwei.hu@bristol.ac.uk

nearly a decade. Second, commercial banks started to and still hold significant amount of excess reserves until now. These new phenomena also brought out significant changes to implementation of monetary policy. In particular, the interest-rate policy is no longer implemented by adjusting the quantity of reserves; instead, it is mainly implemented with interest on excess reserves (IOER).

These new phenomena and new policy measures raise many new research and policy related questions, and this paper is concerned with the following two interconnected ones. On the positive side, is the coexistence of ZLB and large amount of excess reserves a coincidence or is it a result of some common factor? On the normative side, given such coexistence, what is the welfare implications of IOER and what is the optimal IOER? From the perspective of monetary theory, the first question is about coexistence of nominal asset (reserves) and real asset (bank deposits, whose interest payments are presumably backed by real assets), a question related to the coexistence question raised by Hicks (1935). The second question is much debated among policy-makers, and the debate is particularly concerned with the subsiding nature of the IOER and its budgetary implications.[1] The second question is also related to recent debates about negative interest rates. As some economists argued, central bank policy rates can have direct effects on interest rates, and low rates can potentially have unintended consequences to financial intermediation and can result in capital misallocation.[2]

In this paper, we propose a theory of financial intermediation in a monetary model based on Lagos and Wright (2005) aimed at answering these questions. Our model features financial intermediation where banks fund productive assets (which we call *capital*) through deposit issuance. It also features coexistence of deposits and money as means-of-payments, which households acquire to trade in a decentralized market that is plagued by limited commitment and limited record-keeping, the usual frictions that make liquid assets essential. Households can use deposits to trade with formal sellers, who have access to the banking sector, but informal sellers, who have no such access, acquire only cash. Banks in our model are also subject to friction: they cannot commit to repay their deposit obligation and hence, their deposit issuance requires collateral. This implies a pledgability constraint that links banks' lending and borrowing similar to that in Kiyotaki and Moore (1997).

Banks offer deposit contracts in a competitive market. The deposit contract is such that households can withdraw their deposits or deposit their money, depending on their liquidity needs. Banks supply deposits to maximize profits, while households demand deposits and

[1] See, for example, the Economist report ("Is the Federal Reserve giving banks a $12bn subsidy?", printed edition, March 2017) for a description of the public debates on IOER.

[2] For example, Rachel and Summers (2019) argues that "there is a range of concerns about the toxic effects of low rate [...] that they may lead to misallocation of capital by reducing loan payment levels and required rate of return ..."

money for their decentralized trades. Bank profit from funding the capital as it provides returns, but it also profit from deposit issuance when there is a spread between the return on capital and the return promised to deposits. A higher spread induces banks to hold more capital due to the pledgability constraint, and a positive spread will lead to over-accumulation of capital relative to its efficient level determined by its social returns. A this spread decreases with nominal interest rate, our model generates capital misallocation at low rates due to liquidity concerns.

For depositors, this spread corresponds to the cost of holding deposits across periods, and higher spread (and hence lower deposit rate) reduces the deposit demand and lowers decentralized trades. In equilibrium, households hold both money and deposit, and since money can be acquired by forgoing deposits, money demand is inversely related to the nominal interest rate on deposits. Under some mild assumptions, equilibrium in our model has the following features. Majority of money supply is in the form of deposits but currency still circulates, a feature that fits the pattern of money supply in advanced economies. Finally, equilibrium spread is determined by the pledgable value of the productive assets, which in turn is endogenously determined by capital productivity and banks' capital accumulation.

We obtain three main results regarding excess reserves and optimal monetary policies. The first result assumes no interventions, and shows that both excess reserve holding and ZLB occur in equilibrium when productivity is low. The second result demonstrates the welfare effects and the optimality of a restricted class of intervention, IOER financed by nominal asset creation. The third result conducts a mechanism-design exercise and studies the optimal intervention among all incentive compatible and feasible interventions.

We first investigate equilibrium without intervention. The welfare is affected by both capital accumulation and decentralized trades. When the productivity of capital is high and hence pledgable asset is abundant, both the capital market and formal pairwise trades achieve first-best allocations, and the equilibrium spread is zero. That is, capital accumulation is at its efficient level and the interest on deposit is sufficiently high that it is costless to hold deposit and decentralized trade is not constrained by liquidity. The informal trades are inefficient, since high nominal interest rate implies that it is costly to hold money.

In contrast, when productivity is low and hence collateral is scarce, the equilibrium spread is positive and hence interest on deposits is low, there is both overaccumulation of capital and inefficient decentralized trade in the formal sector. Banks do not hold reserve in equilibrium, however, until the productivity is sufficiently low and the economy features ZLB. The equilibrium interest rate on deposits can never go below zero, as the rate-of-return on reserve sets a floor for interest rate on deposits due to no arbitrage. Indeed, if rates are negative, banks can hold reserve and issue deposits against it to make profits. Therefore, when productivity is sufficiently low, the interest on deposits will hit the ZLB and stay there. Moreover, banks hold excess reserves in equilibrium to bridge the gap of liquidity demanded by the households at ZLB. Finally, at ZLB, since there is no opportunity cost of holding money, trades in informal meetings are at their first-best level.

Our first main result then shows that, without interventions, banks hold excess reserve in equilibrium if and only if there is ZLB. They also point to low productivity as the underlying cause, and there is a spillover effect to decentralized trades when productivity is low. In our model, the same situation can be caused by more severe financial frictions that reduce the pledgability of the productive assets.

The two other results characterize optimal monetary policy with interest on reserves and optimal liquidity requirements. We first consider interventions that pay interest on reserves financed by creation of nominal assets. We show that it is optimal to have a positive IOER, which requires a positive inflation rate as well, when the productivity is close to or below the threshold for ZLB. In this circumstance, paying interest on excess reserve helps both liquidity provision for households and the alleviation of the over-accumulation. IOER incentivise the banks to pay higher interest on their deposits and increase the real value of the reserves,[3] which also decreases capital accumulation. This benefits decentralized trades using deposits and efficiency in capital accumulation.

This result is related to earlier ideas that IOER can implement the Friedman (1959) rule.[4] Indeed, if one can finance an interest on reserves that is equal to the discount rate, then both the capital market and the formal decentralized trades would have first-best allocations. However, it also has a redistributional effects between deposit users and cash users, and imposes an implicit tax on informal trades which decreases their volume. The optimal policy trades off these two effects. When ZLB or very low interest nominal rate occurs, trades using cash are at or close to the first-best level and hence distorting them is marginally costless. As a result, a positive inflation is optimal to subsidise interest on reserves to alleviate inefficiency in capital over-accumulation and formal trades.

Moreover, if price stability is imposed, i.e., a targeted inflation rate is given, optimal interest rate is procyclical. We illustrate this result in Figure 1 below with inflation target at $\bar{\pi}$, where the left panel depicts the nominal interest rate on deposits without any intervention (other than the targeted inflation rate) and the optimal IOER, and the right panel depicts the corresponding equilibrium reserve holdings. The optimal IOER eventually coincides with the nominal rate without intervention as productivity recovers, but a remarkale feature of

[3] Keister (2016) argues that this transmission from IOER to interest on deposits is a major benefit of IOER, and points that the transmission is not one-for-one due to various costs of deposit issuance. We can also introduce those costs but that would not affect the main results.

[4] See, for example, Goodfriend (2002) for a related argument.

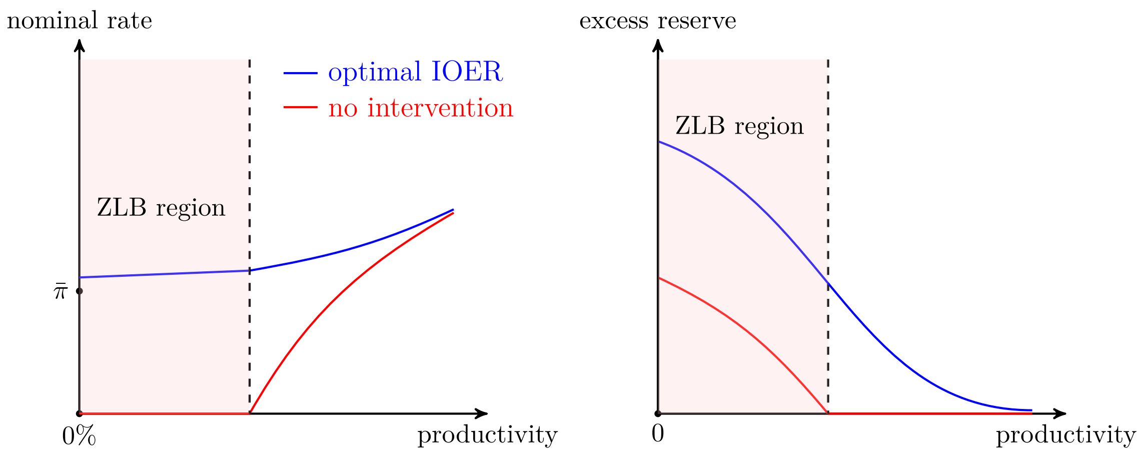

Figure 1: Interest rate and excess reserves under no intervention and optimal IOER

the optimal policy is that optimal IOER is always positive and, when the number of informal sellers is small, is above the targeted inflation rate and optimal equilibrium excess reserve holding can still be large even for productivity levels above the ZLB region.

Our third main result characterizes optimal monetary policy among all incentive compatible and fiscally feasible policies. We consider a consolidated government budget constraint, and allow for explicit taxation or implicit taxation (through inflation or liquidity requirements) that respect voluntary participation. Consistent with the underlying frictions that render money and deposits essential for their transaction roles, households can be taxed through the banking system, the only place where they are recorded. Moreover, banks can opt out as private equity and issue no deposits to avoid regulation and taxation.[5] We obtain a full characterization of the constrained-efficient allocations, as well as the optimal interventions to implement them. When the productivity is sufficiently high, then the first-best allocations in capital and both decentralized trades are implementable. Otherwise, the constrained efficient allocation features both inefficient pairwise trades and capital over-accumulation, and both bank reserves and a positive IOER are essential for welfare.

Regarding the optimal intervention, there is a threshold of productivity below which optimal inflation rate is positive. When that is the case, the optimal policy imposes a fixed liquidity requirement and a proportional liquidity requirement to asset holdings. More precisely, these regulations require banks to hold a fixed amount plus a given proportion (relative to their asset holdings) of nominal assets,[6] regulations that are conceptually similar to the liquidity requirements in Basel III regulations. Since interest is not paid on required

[5]The assumption here is that banks become public entity only if they issue deposits. One can assume that banks are always public and their income always subject to taxation; this will not alter our main results.

[6]In our model these are reserves. Of course, we could have introduced nominal government bonds and required the banks to hold such assets. However, as in Williamson (2012), this would not make any difference.

reserves, these requirements imply implicit taxes. The fixed liquidity requirement is useful as it provides funding through taxing excess profits banks have above their outside option as private equity. The proportional requirement is designed to correct a pecuniary externality from deposit issuance–banks are tempted to issue more deposits when interest on deposits are low, which in turn causes more capital over-accumulation. In contrast, for productivity levels above that threshold, it is optimal to impose both a fixed fee and a proportional tax on asset holdings to the banking sector, and to have deflation.

This third result implies that, except possibly for very high productivity levels, intervention is always essential to implement the constrained-efficient allocation, a result reminiscent of Wallace (2014) for pure-currency economies. It also shows that IOER and excess reserves, together with other complimentary measures identified here, are indeed optimal monetary policy. Finally, when inflation is optimal, the optimal real interest on reserves is increasing in productivity. The behaviour of the optimal inflation rate depends on how informal trades enter the welfare function. It decreases with productivity if it causes no externality. In contrast, if informal trades exhibit negative externality, as some have argued,[7] then the optimal inflation rate may in fact increase with productivity and hence optimal nominal interest rate increases as well.

Related literature

We are not the first to study ZLB in a Lagos-Wright (2005) model. Williamson (2012) also has a model where both money and deposits circulate and shows that ZLB occurs when assets used as collateral are scarce. Different from that paper, we consider endogenous supply of capital, introduce frictions in the banking sector, and focus on optimal liquidity requirements. Our ZLB result is also related to the rate-of-return equality result in Lagos and Rocheteau (2008), where both money and capital can be used as means-of-payments and monetary equilibrium exist only if capital earns a zero net return. Hu and Rocheteau (2013) also adopts a mechanism-design approach in an environment where both money and capital can serve as means-of-payments. As here, it is also shown there that nominal assets are essential to implement the constrained efficient allocation. Finally, we are the first to consider banking regulation as part of monetary policy. Stein (2012) also considers monetary policy in the context of financial intermediation, and points out an externality of deposit issuance according to which banks issue too much short-term debt. Different from there, however, the main distortion here is capital over-accumulation.

[7]See, for example, Sands (2016) and Rogoff (2016).

2 Environment

The baseline model is borrowed from Rocheteau and Wright (2005), and we add capital goods in a fashion similar to that in Lagos and Rocheteau (2008). Time is discrete and has an infinite horizon, $t \in \mathbb{N}_0$. The economy is populated by three sets of agents. The first set consists of a unit measure of *consumers*, the second consists of measure $\sigma < 1$ of *formal sellers* and measure $\sigma_u < 1 - \sigma$ of *informal* or *underground sellers*, and the third consists of measure n of *firms*. As will be explained below, only formal sellers have access to the banking sector while informal sellers do not.

There is an infinite time horizon and each period has two stages. The first stage features random pairwise meetings between consumers and sellers in a decentralized market (DM). In the DM, consumers are randomly matched with formal or informal sellers in pairs: the probability of matching a formal seller is σ and that of matching an informal one is σ_u. Each consumer can have at most one meeting in each DM. The second stage has a centralized market (CM) where all agents meet. The CM good is taken as the numéraire.

Both consumers and firms have access to a linear technology to produce the CM good: with h_t units of labour input $x_t = h_t$ units of CM good can be produced. The firms, in addition, specialize in CM production with *capital*. The CM good can be transformed into capital good one-for-one, but only firms have the expertise to obtain returns from them. To do so, it requires labour input from the firms to operate capital; to operate on k units of capital, it requires $\psi_0(k)$ units of labour. Thus, the total cost for a firm to operate k units of capital is equal to $\psi(k) \equiv \psi_0(k) + k$. By holding k units, the capital will produce Ak units of CM goods in the following CM, where A measures capital *productivity*. We assume that ψ_0 is twice differentiable with $\psi_0'(0) = 0 = \psi_0(0)$, $\psi_0'(k) > 0$, and $\psi_0''(k) > 0$. We assume that capital is fully depreciated after one period.[8]

A consumer's preference are represented by the expected discounted sum of utilities,

$$-h_0 + x_0 + \mathbb{E} \sum_{t=1}^{\infty} \beta^t \left[u(q_t) + v(y_t) + x_t - h_t \right],$$

where $\beta \equiv (1 + r)^{-1} \in (0, 1)$ is the discount factor, q_t is the period-t DM consumption produced by a formal seller and y_t by an informal seller (note that $q_t \cdot y_t$ for each t), x_t is CM consumption, and h_t is the supply of hours in the period-t CM. A formal seller's preference

[8] The assumption that the return of capital is linear can be relaxed to concave returns without affecting the main results but with significantly more complicated algebra. The assumption of full depreciation is with no loss of generality under putty-putty capital–with partial depreciation we can simply add unappreciated capital into the returns.

and an informal seller's preferences are, respectively,

$$\sum_{t=1}^{\infty} \beta^t \left(-q_t + x_t\right) \text{ and } \sum_{t=1}^{\infty} \beta^t \left(-y_t + x_t\right),$$

where q_t and y_t are their respective period-t DM production and x_t period-t CM consumption. The DM utility functions, $u(q)$ and $v(y)$, are both strictly increasing and concave, twice differentiable, and $u(0) = 0 = v(0)$, with $q^* = \arg\max\left[u(q) - q\right]$ and $y^* = \arg\max\left[v(y) - y\right]$. Moreover, $u'(0) = \infty = v'(y)$.

Consumers cannot commit to their future actions and have no record-keeping technology on their own, and hence cannot promise repayment for their consumption in DM meetings. Instead, they need to acquire liquid assets to finance their DM consumption. We consider both private and public liquidity provision. For private liquidity provision, the firms can choose to become *banks* by issuing bonds to finance their capital holdings, and consumers can acquire them in the CM and transfer them to formal sellers in DM. For public liquidity provision, the government issues two nominal assets. The first consists of fiat money that can be used as means-of-payment for consumers. The second consists of reserves that can be held by banks, which, as accounts in the central bank, are fully pledgable to back deposit issuance. There is perfect convertibility between money and reserves in the CM and they are traded at the same price.[9]

As will become clear later, all firms choose to become banks in equilibrium and we refer to them as banks, reflecting the fact that their liabilities are circulated as means-of-payment. However, only formal sellers have access to the banking system and accept deposit transfers from consumers, while informal sellers only accept money. This lack of access can be justified either by technological feasibility for certain trades, or it can be due to the illegal nature of products the informal sellers produce, such as illicit drugs.[10] These trading opportunities make money essential in the sense of Wallace (2001).

There is a public record of banks' liabilities and capital holdings, but banks have limited liability and cannot commit to their future actions. If a bank files for bankruptcy, the court could seize up to ρ proportion of his capital income. Thus, by holding k units of capital, a bank can credibly pledge up to ρ fraction of the returns from capital that he invested in but can take the rest away, a friction similar to that in Kiyotaki and Moore (1997). Banks maximize their life-time profits with discount factor β.

Each unit of the bond is a promise to pay one unit of coming CM good with its price

[9]One can imagine a system where the government separate these two assets and set the exchange rate between them as a policy variable, but this would not change our results here.

[10]Some economists argue that such trades would cause negative externality to others and we consider this in our concluding discussions.

denoted by φ. Alternatively, one can interpret this contract as a *deposit contract*: the household "deposits" φ units of capital in current CM and is promised one unit of CM goods next period. The implied (real) interest rate for the deposit contract is then

$$r_d \equiv 1/\varphi - 1. \tag{1}$$

In the CM there is a competitive market for money, deposit, reserve, and CM good.

In addition to the competitive market in the CM, we assume that at the beginning of the DM, consumers receive some information about their trading opportunity and can rebalance their portfolio. In particular, they first learn that whether they will have an informal meeting,[11] followed by a competitive market where consumers trade deposits for money or vice versa. This captures an essential feature of deposit contract–deposits can be converted to currency when needs arise. This market corresponds to the ATM machines where people can save/withdraw currency to/from their bank accounts. We use χ to denote the relative price between money and deposit in this market and refer to this market as the DM money/deposit market.

We are interested in the case where consumers have sufficient deposits to be withdrawn in equilibrium. For that purpose, we assume that the liquidity needs in the underground meetings is relatively small:

$$v(y) = \theta u(y) \text{ with } \theta < \frac{\sigma}{\sigma + r}. \tag{2}$$

Finally, we define social welfare in our economy, and we consider stationary allocations only. An allocation consists of DM trade per formal and informal meeting, denoted by q and by y, respectively, and the amount of capital held by each bank, k, which are constant across time periods. Given an allocation (q, y, k), the total welfare is given by

$$\mathcal{W}(q, y, k) = \underbrace{\sigma[u(q) - q]}_{(a)} + \underbrace{\sigma_u\left[v(y) - y\right]}_{(b)} + \underbrace{n\left[Ak - (1 + r)\psi(k)\right]}_{(c)}, \tag{3}$$

in which terms (a) and (b) measure efficiency in DM production, and term (c) measures efficiency in capital production. The first-best allocation, defined as the allocation (q, y, k) that maximizes $\mathcal{W}(q, y, k)$, is denoted by (q^*, y^*, k^*). The allocation q^* maximizes term (a), y^* maximizes (b), and k^* maximizes (c) and satisfy the following FOC's: $u'(q^*) = 1$, $v'(y^*) = 1$, and $\beta A \leq \psi'(k^*)$ with equality if $k^* > 0$. We use $k^*(A)$ to denote the first-best

[11]We can also assume that the households also learn if they would have a pairwise meeting with other households; this would not change any qualitative results.

capital holding under A. Note that $k^*(A) = 0$ if and only if $A \leq 1 + r$ and $k^*(A)$ strictly increases with A for $A > 1 + r$. To simplify exposition, we assume $A > 1 + r$ from now on.

We remark that $k^*(A)$ can be implemented if firms cannot issue bonds/deposits. In this case, a firm chooses his capital holding to maximize

$$\Pi = -\psi(k) + \beta Ak.$$

Note that the opportunity cost cost of holding k capital includes both the CM good converted, k, plus the labour cost, $\psi_0(k) = \psi(k) - k$, while the return Ak only occurs in the next CM and hence is discounted. The FOC is given by $\psi'(k) = \beta A$, or, $A = (1 + r)\psi'(k)$, and hence this implements $k = k^*(A)$. Finally, we describe the time line and the general characteristics of the banking contracts.

The course of events. In the period-t DM, the course of events is as follows:

1. first, consumers learn whether they will have informal meetings or not, and then readjust their portfolio in the DM money/deposit market, with relative price χ (for money in terms of deposits);

2. this is followed by pairwise meetings:

    - a consumer with a formal meeting makes a take-it-or-leave-it offer, (q, d, p), to the seller, where q denotes production, d denotes deposit transfer, and p denotes money transfer;

    - a consumer with an informal meeting makes a take-it-or-leave-it offer, (y, p), to the seller, where y denotes production, and p denotes money transfer.

In the period-t CM, the course of events is as follows:

1. first, banks receive output from their capital holdings and settle obligations with those who hold deposits;

2. then, banks decide their new capital holdings and reserve holdings;

3. banks issue deposit contract, one unit sold at price φ to consumers.

Note that since we focus on stationary equilibrium only, prices are constant over time.

3 Equilibrium analysis and ZLB

Now we turn to equilibrium analysis, beginning with a bank's problem. Given φ, a bank's profit by holding k capital, z reserves, and issuing b deposits at period-t CM is given by

$$\begin{aligned} \Pi(k,b,z;\varphi) &= b\varphi - z - \psi(k) + \beta(Ak + z - b) = \beta\left\{sb + Ak - rz - (1+r)\psi(k)\right\}, \quad (4) \\ \text{where} \quad & s \equiv (1+r)\varphi - 1. \end{aligned}$$

According to (4), by issuing b units of deposits and holding k units of capital, the bank needs to provide by itself $k - \varphi b$ units of CM good (negative means consumption) to convert to the capital goods plus the labour cost of $\psi(k) - k$ in the current period, and will receive Ak units of CM good from the capital holding but needs to repay b to depositors in the coming CM, while the reserves (measured in terms of the CM good) pay no net interest. The variable s will play a prominent role in our analysis. For banks, it measures the spread between the funds obtained from deposit issuance and the return they promise to pay to depositors, taking discounting into account; a positive s means a positive margin for banks' profits by issuing deposits. As we will see below, s also measures the cost of holding deposits for consumers. Two levels of s are of particular interest: when $\varphi = 1$ and $r_d = 0$ (recall (1)), $s = r$, and this implies a zero interest rate; when $\varphi = \beta$ and $r_d = r$, $s = 0$, and this sets a ceiling for the deposit rate, above which banks make negative profits from issuing deposits.

A bank chooses k, b and z to maximize (4) subject to the pledgeability constraint,

$$b \leq \rho A k + z, \tag{5}$$

which limits a bank's amount of deposit issuance according to his capital and reserve holding. As mentioned, a bank can only credibly pledge ρ fraction of capital income while reserves are fully pledgable. In this sense, the bank deposits are collateralized liabilities. Note that whenever $s > 0$, the constraint (5) is binding, that is, firms are better off to become banks. Substitute b by $\rho A k + z$, the profit function (4) becomes

$$\Pi(k,b,z;\varphi) = \beta\left\{(1+s\rho)Ak - (r-s)z - (1+r)\psi(k)\right\}, \tag{6}$$

and the optimal capital holding, $K(s, A)$, is determined by the following FOC:

$$(1+s\rho)A = (1+r)\psi'(k). \tag{7}$$

Note that the solution works for $s = 0$ as well. For reserve holdings, (6) implies that equilibrium $s \leq r$, for otherwise there will be infinite demand on reserves. It also implies that banks hold reserve in equilibrium only if $s = r$. Because of the pledgability constraint, a bank needs to produce in the CM to finance part of his capital holding. Thus, the pledgeability constraint works as a proportional capital requirement imposed by market discipline.

Now we turn to consumer behaviour. In the beginning of the DM, we call a consumer an *informal* consumer if he learns that he will meet an informal seller, and we call him a *formal* consumer otherwise. We use $V(b, m)$ to denote a consumer's continuation value upon entering the DM with deposit b and money m, and use $W(b, m)$ to denote a consumer's continuation value upon entering the CM with deposit b and money m. Standard Lagos-Wright (2005) arguments show that $W(b, m) = b + m + W(0, 0)$, that is, W is linear in b and in m. Denote $W^0 \equiv W(0, 0)$.

We solve the problem for formal and informal consumers in the DM money/deposit market, with χ as the price for money (in term of deposits). Now, if $\chi < 1$, no one will hold deposit. Hence, we assume $\chi \geq 1$. The opportunity cost of obtaining money is then $\chi - 1$, which, as we show later with a no-arbitrage argument, in equilibrium is equal to r_d. Consider now an informal consumer with portfolio (b, m). His continuation value is then (where (b^i, m^i) is his portfolio after the DM money/deposit market)

$$V^i(b, m) = \max_{(b^i, m^i, y)} v(y) + b^i + (m^i - y) + W^0,$$

subject to $b^i + \chi m^i \leq b + \chi m$ and $m^i - y \geq 0$. When $\chi > 1$, the informal consumer will only hold just enough money to spend in the coming meeting and hence choose $y = m^i$; when $\chi = 1$, he is indifferent between money and deposit. In both cases, we can reduce the problem to (where $w^i = b^i + \chi(m^i - y)$)

$$V^i(b, m) = \max_{w^i \geq 0} v\left[\frac{1}{\chi}(b - w^i) + m\right] + w^i + W^0, \tag{8}$$

with the solution given by $b^i = b + \chi(m - m^i)$ and $m^i \geq y$ (with equality whenever $\chi > 1$), where y is determined by the FOC

$$v'(y) = \chi \text{ if } y \leq \frac{1}{\chi}b + m \text{ and } y = \frac{1}{\chi}b + m \text{ otherwise.} \tag{9}$$

It is then straightforward to verify that

$$\frac{\partial}{\partial b}V^i(b, m) = v'(y)\frac{1}{\chi},\ \frac{\partial}{\partial m}V^i(b, m) = v'(y). \tag{10}$$

Similarly, for a formal consumer with portfolio, (b, m), entering the DM, his portfolio after the DM money/deposit market, denoted by (b^f, m^f), will be such that $b^f = b + \chi m$ and $m^f = 0$ if $\chi > 1$ and he is indifferent between money and deposit if $\chi = 1$. In both cases, his continuation value is (note that the probability that a formal consumer has a successful meeting with a formal seller is $\sigma/(1-\sigma_u)$)

$$V^f(b,m) = \max_{q \le b+\chi m} \frac{\sigma}{1-\sigma_u}[u(q) - q] + b + \chi m. \tag{11}$$

The optimal solution is $q = q^*$ if $q^* \le b + \chi m$, and $q = b + \chi m$ otherwise. Thus,

$$\frac{\partial}{\partial b}V^f = \frac{\sigma}{1-\sigma_u}[u'(q) - 1] + 1, \ \frac{\partial}{\partial m}V^f = \frac{\sigma}{1-\sigma_u}\chi[u'(q) - 1] + \chi. \tag{12}$$

Thus, the continuation value upon entering the DM with portfolio (b, m) is given by

$$V(b,m) = \sigma_u V^i(b,m) + (1-\sigma_u)V^f(b,m),$$

and hence the CM problem is $\max_{b,m} -\varphi b - m + \beta V(b,m)$. The FOC's are then given by

$$-\frac{\varphi}{\beta} + \sigma_u \frac{1}{\chi}[v'(y) - \chi] + \sigma[u'(q) - 1] + 1 = 0, \tag{13}$$
$$-\frac{1}{\beta} + \sigma_u[v'(y) - \chi] + \chi\sigma[u'(q) - 1] + \chi = 0. \tag{14}$$

These FOC's imply that $\chi = 1/\varphi$, for otherwise they cannot be satisfied simultaneously. As we have seen from bank optimization, in equilibrium $s \le r$ and hence $\varphi \le 1$, which in turn verifies that $\chi \ge 1$. Following the same logic, we can show that $\chi < 1$ cannot be an equilibrium. Since $\chi = 1/\varphi$ and since consumers can always rebalance his holdings in the DM money/deposit market, they are indifferent between money and deposit in the CM. However, the demands for deposits and money in the CM are determined by those demands in the DM money/deposit market and market clearance there:

$$(1-\sigma_u)b^f + \sigma_u b^i = b, \ (1-\sigma_u)m^f + \sigma_u m^i = m. \tag{15}$$

This condition also highlights a crucial difference from the CM market–the total amounts of deposit and real balances in the DM money/deposit market are constrained by the amount brought from the previous CM market.

The following lemma gives sufficient conditions for optimal y to be determined by the first case in (9), i.e.,

$$v'(y) = \chi = 1/\varphi = (1+r)/(1+s), \tag{16}$$

and the solution is denoted by $Y(s)$. Moreover, in this case, (13) is reduced to

$$-s + \sigma[u'(q) - 1] + 1 = 0, \tag{17}$$

Let $Q(s)$ denote the solution to (17). Note that for any $s > 0$, $Q(s)$ is uniquely determined; when $s = 0$, $Q(s)$ is not pinned down but $Q(s) \geq q^*$. With no loss of generality we define $Q(0) = q^*$. Then, $Q(s)$ is continuous and strictly decreasing in s for all $s \geq 0$.

Lemma 3.1. *Suppose that $v'(y)y$ is increasing and that (2) holds. Then, for $s \in [0, r)$, the demand for money is given by $M(s) = \sigma_u Y(s)$, and deposit demand given by*

$$D(s) = Q(s) - \chi M(s) = Q(s) - \frac{1+r}{1+s}\sigma_u Y(s). \tag{18}$$

When $s = r$, optimal $q = Q(r)$ and optimal $y = Y(r) = y^*$.

When $s < r$ and hence $\chi < 1$, Lemma 3.1 shows that the demand for money, $M(s)$, increases with s and hence decreases with r_d. In (16), the marginal benefit of holding one extra unit of money into an informal meeting is $v'(y) - 1$, which is equal to the opportunity cost of holding money, $\chi - 1 = r_d$. When $s < r$ and $r_d > 0$, this implies that only informal consumers demand real balances and market clearing determines $M(s)$, that is, $m = M(s) = \sigma_u Y(s)$ and $b = D(s)$ given by (18) satisfy (15) with $m^i = Y(s)$ and $m^f = 0$. Given s, bank pledgability constraint (5) and optimization (7) imply that the market-clearing condition for deposits is given by (where z is reserve holding per bank)

$$Q(s) - \frac{1+r}{1+s}\sigma_u Y(s) = D(s) \quad \leq \quad n\rho A K(s, A) + nz, \tag{19}$$

with equality whenever $s > 0$. When $s = r$ and $r_d = 0$, there is no opportunity cost of holding money and hence consumers can demand more than $\sigma_u Y(r)$ in the CM. The condition (19) is still sufficient for market clearing but not necessary. The following proposition fully characterizes equilibrium allocations and interest rates.

Proposition 3.1. *Assume (2) and that $v'(y)y$ is increasing. There is a monetary equilibrium characterized by two thresholds, $\bar{A} < A^*$, that solve*

$$\begin{aligned} q^* &= n\rho A^* k^*(A^*) + \sigma_u(1+r)Y(0); && (20) \\ Q(r) &= n\rho \bar{A} K(r, \bar{A}) + \sigma_u Y(r). && (21) \end{aligned}$$

(1) For $A \geq A^*$, equilibrium is unique with $q = q^*$, $k = k^*(A)$, $y < y^*$, $z = 0$, and $s = 0$.

(2) For $A \in [\bar{A}, A^*)$, equilibrium is unique with $s = s(A) \in (0, r]$, $q < q^*$, $k > k^*(A)$, and $y < y^*$. Equilibrium $s(A)$ and y strictly decrease, and q strictly increases with A.

(3) For $A < \bar{A}$, equilibrium $s = s(A) = r$, $q = Q(r) < q^*$, $k = K(r, A) > k^*(A)$, and $y = y^*$. Equilibrium z is indeterminate but satisfies

$$0 \leq z \leq \frac{1}{n} \left[Q(r) - \sigma_u Y(r) - n\rho A K(r, A) \right], \tag{22}$$

where the upper bound strictly decreases with A.

Proposition 3.1 (1) and (2) show that when $A \geq \bar{A}$, banks do not hold reserves in equilibrium. Moreover, when A is sufficiently large ($A \geq A^*$), the equilibrium allocation achieves first-best level of formal trades in the DM and for capital holdings, notably without any interventions, but informal trades are inefficient. For lower A's, formal trades become inefficiently low, and capital accumulation is above its first-best level. Moreover, equilibrium spread $s(A)$ increases as A decreases and hence real deposit rate, r_d, decreases as capital becomes less productive. By (16), equilibrium y increases as s increases, and it reaches the first-best level, y^*, when $s = r$. These features match the basic patterns of money supply in advanced economies: currency is only a small fraction of the money supply while deposit is the major component, and the demand for currency decreases with nominal interest rate on deposits. We depict the determination of $s(A)$ for $A \in (\bar{A}, A^*)$ in the left panel of Figure 2 below.

In contrast, Proposition 3.1 (3) shows that when $A < \bar{A}$, banks hold reserves in equilibrium and the equilibrium rate-of-return on deposits, r_d, is zero by $s = r$ and (1), featuring the zero-lower-bound. Reserves set a floor on returns to deposits: a negative interest rate on deposits would induce banks to hold more reserves (which have zero interest) and issue deposits to arbitrage, which cannot occur in equilibrium. Instead, banks compete by holding reserves until the return to deposits is exactly zero. In the right panel of Figure 2 below, we depict how the largest equilibrium reserve holding is determined. The motive for banks to hold reserves is independent of informal trades. In particular, the threshold $\bar{A}$ is determined in the same way when $\sigma_u = 0$ and hence reserves are still essential even if real assets can be used (indirectly) to provide liquidity in all meetings. The result that largest equilibrium reserves increases as A decreases shows that nominal assets become more important as real assets become more scarce. Note that this does not contradict the quantity theory of money, as we are comparing steady-states across different A's, and the price for money is constant in such an equilibrium even when there is excess reserves because of the *demand* from banks for reserves, a similar point also made by Williamson (2012). As discussed in concluding

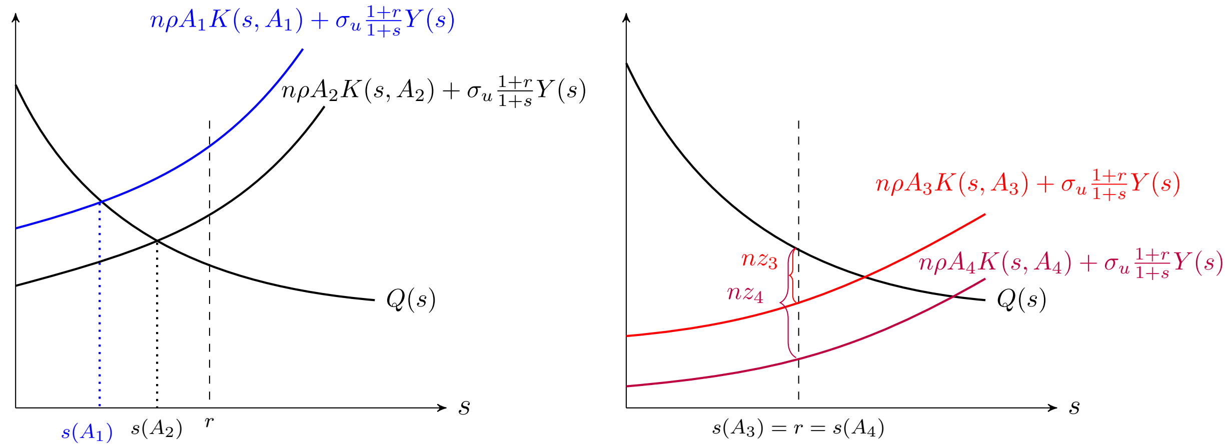

Figure 2: Determination of s (left) and excess reserves (right) ($A_1 > A_2 > A3 > A_4$)

remarks, our model would predict deflation if we introduce transition dynamics when ZLB occurs.

When ZLB occurs, monetary equilibrium may not be unique, as consumers can demand more money than what is needed for informal trades, but equilibrium interest on deposits and equilibrium capital holdings are uniquely determined. The total value of nominal assets (reserve plus money) is uniquely determined, and an arbitrarily small interest on reserves will select the equilibrium with the upper bound on equilibrium reserves in (22). Besides restoring a unique equilibrium, when ZLB occurs, a constant money/reserve supply does not seem the best policy and we turn to optimal interest on reserves in the following sections.

4 Inflation and optimal interest on excess reserves

Here we consider interest on reserves.[12] We assume that the central bank finance such interests by reserve/money creation. We consider stationary schemes in which the net creation rate of total nominal assets is constant. We use M_t to denote the average amount of reserves and money per consumer at the end of period t, and use ϕ_t to denote the price for both reserves and money (in terms of numéraire) at period-t CM.[13] For policy parameters, we denote the constant net money/reserve creation rate by π, and the (nominal) interest rate on excess reserves (IOER) by ι, and the pair (ι, π) characterizes the central bank policy. We use R_t to denote average reserve holding, and C_t average money holding, both in per

[12] Since we do not impose required reserves, all reserve holdings here are excess reserves.

[13] Note that since there is perfect convertibility between money and reserve in the CM, the price must be the same for the two assets.

consumer terms. Central-bank budget balancedness requires

$$\iota R_t \leq \pi M_t = \pi(R_t + C_t). \tag{23}$$

In (23), we have an inequality and the central bank does not have to use all the seigniorage revenue; whatever left is redistributed to consumers in a lump-sum fashion.

Stationarity implies that $M_t \phi_t$ is a constant over time and hence $\phi_t = \phi_{t+1}(1+\pi)$, and this allows us to express nominal assets and policy parameters in real terms. Let $\hat{m} \equiv C_t \phi_{t+1}$ denote the real value of average money holding (per consumer) in terms of next-period CM good, and $\hat{z} \equiv R_t(1+\iota)\phi_{t+1}$ the real value of average reserve holding (per consumer) in the coming CM (after the interest). Thus, we may rewrite (23) as

$$\frac{\iota}{1+\iota}\hat{z} \leq \pi\left(\hat{m} + \frac{1}{1+\iota}\hat{z}\right). \tag{24}$$

Now we turn to bank problem. Given a policy, (π, ι), and φ, a bank chooses capital holding, k, reserve holding, z, and deposit issuance, b, at period-t CM to maximize

$$\begin{aligned}\Pi(k, b, z; \varphi, A) &= b\varphi - \frac{1+\pi}{1+\iota}z - \psi(k) + \beta\{Ak + z - b\} \\ &= \beta\left\{sb + Ak - (1+r)\psi(k) - \frac{\iota_m - \iota}{1+\iota}z\right\},\end{aligned}$$

where $\iota_m \equiv (1+\pi)/\beta - 1$, subject to the pledgeability constraint (5). Note that z is measured in term of period-$(t+1)$ CM good. Substituting b by taking (5) at equality, we have

$$\Pi(k, b, z; \varphi, A) = \beta\left\{(1+\rho s)Ak - (1+r)\psi(k) - \left(\frac{\iota_m - \iota}{1+\iota} - s\right)z\right\}. \tag{25}$$

If $s > (\iota_m - \iota)/(1+\iota)$, there will be infinite reserve demand and this cannot be an equilibrium. If $s < (\iota_m - \iota)/(1+\iota)$, no banks hold reserves in equilibrium. Thus, banks hold reserves in equilibrium only if the cost of holding them is exactly zero, that is, only if

$$s = \frac{\iota_m - \iota}{1+\iota}. \tag{26}$$

Note that (26) holds if and only if $\iota = \iota_d \equiv (1+\pi)/\varphi - 1$, the nominal interest rate paid on deposits. That is, the IOER is passed on to the consumers. Whenever $s \leq (\iota_m - \iota)/(1+\iota)$, the optimal capital holding is still determined by (7).

For the consumer's problem, inflation only affects (14), which is now changed to

$$-\frac{1+\pi}{\beta}+\sigma_u[v'(y)-\chi]+\chi\sigma[u'(q)-1]+\chi=0.$$

This then implies that $\chi=(1+\pi)/\varphi=(1+\iota_m)/(1+s)$. Under assumption (2), the equilibrium q is still given by $Q(s)$. The demand for money is determined by $M(\iota_m,s)/\sigma_u=Y(\iota_m,s)$, where $Y(\iota_m,s)$ solves the FOC (16) but with r replaced by ι_m. The following lemma characterizes implementable equilibrium allocation.

Lemma 4.1. *Assume (2) and that $v'(y)y$ is increasing. For any $A<A^*$, a policy (π,ι) and s satisfying (26) are implementable if and only if $s\in[\underline{s}(\iota_m,A),\overline{s}(\iota_m,A)]$, where $\overline{s}(\iota_m,A)$ is the highest $s\le\iota_m$ that satisfies the first inequality below, and $\underline{s}(\iota_m,A)$ is the lowest $s\ge 0$ that satisfies the second inequality below:*

$$n\rho AK(s,A)\le Q(s)-\frac{1+\iota_m}{1+s}\sigma_uY(\iota_m,s)\le n\rho AK(s,A)+\frac{\iota_m-r}{r-s}\sigma_uY(\iota_m,s). \tag{27}$$

The assumptions in Lemma 4.1 ensures that for any policy (π,ι) and $s\in(0,\iota_m]$, the deposit demand is given by $D(\iota_m,s)=Q(s)-\frac{1+\iota_m}{1+s}\sigma_uY(\iota_m,s)$ and the money demand is $M(\iota_m,s)=\sigma_uY(\iota_m,s)$. Accordingly, the second inequality in (27) considers the case where all the seigniorage revenue is used to pay interest on reserves and, by (24), the maximum value of equilibrium reserves implementable is given by $\frac{\iota_m-r}{r-s}\sigma_uY(\iota_m,s)$. Since left-side is strictly decreasing in s and the right-side is strictly increasing, the second inequality gives a lower bound on s. Since $\frac{\iota_m-r}{r-s}\sigma_uY(\iota_m,s)$ goes to infinity as s approaches r from below, this lower bound is always smaller than r. See also Figure 3 for the determination of $\underline{s}(\iota_m,A)$.

In contrast, the first inequality in (27) considers the case where all seigniorage revenue is paid back to consumers lump-sum, and hence it corresponds to (19) for the given ι_m. Since $(1+\iota_m)Y(\iota_m,s)$ strictly decreases with ι_m, $\overline{s}(\iota_m,A)>s(A)$, with $s(A)$ from Proposition 3.1. When the highest s satisfying that inequality is higher than ι_m and hence $\overline{s}(\iota_m,A)=\iota_m$ (see the right panel of Figure 3 for such a situation), banks hold excess reserves in equilibrium and nominal interest rate is zero. That is, the ZLB result extends to lump-sum inflation as well. Otherwise, $\overline{s}(\iota_m,A)$ can be implemented with ι that satisfies (26), but banks hold no excess reserves in equilibrium (see the left panel of Figure 3 for such a situation).

Three remarks regarding the bounds are in order. First, as ι_m converges to r, both $\underline{s}(\iota_m,A)$ and $\overline{s}(\iota_m,A)$ converge to $s(A)$. In particular, if $A<\bar{A}$, $\underline{s}(\iota_m,A)$ converges to r as ι_m converges to r and $\sigma_u\frac{\iota_m-r}{r-s}Y(\iota_m,s)$ converges to the largest equilibrium nz in Proposition 3.1 (3). Second, for any given implementable policy with $s<\iota_m$, the monetary equilibrium is unique. Third, since $K(s,A)$ is increasing in A, it is easy to see from Figure 3 that both

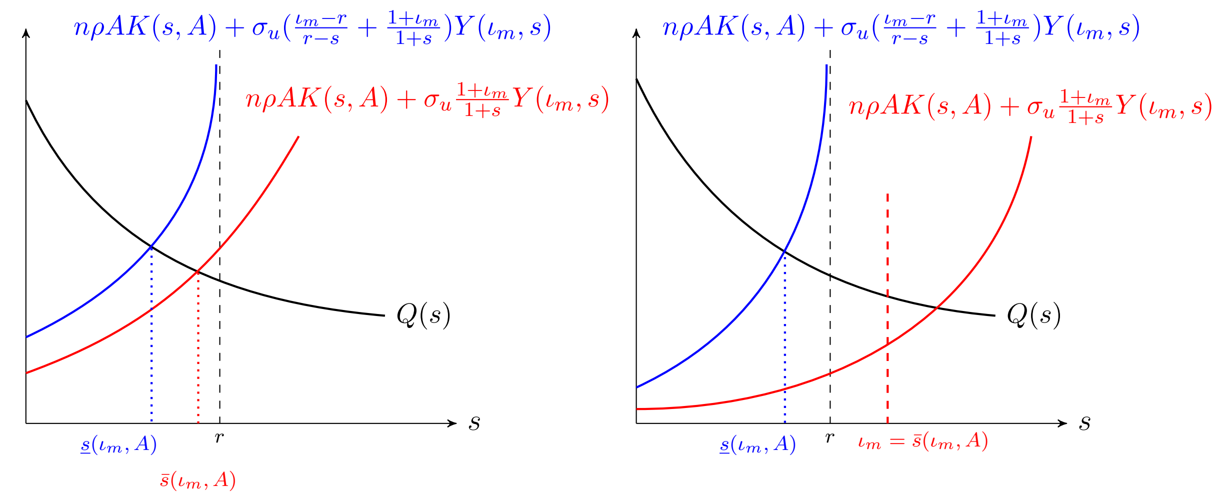

Figure 3: Determination of $\underline{s}(\iota_m,A)$ and $\bar{s}(\iota_m,A)$

$\bar{s}(\bar{\iota}_m,A)$ and $\underline{s}(\bar{\iota}_m,A)$ are decreasing in A.

Note that we consider $A<A^*$ only. When $A\geq A^*$, one can only implement $s=0$, and it is optimal to set $\iota_m=r$, i.e., $\pi=0$, and set $\iota=r$. For $A<A^*$, Lemma 4.1 implies that the central bank then maximizes (3) subject to $s\in[\underline{s}(\iota_m,A),\overline{s}(\iota_m,A)]$. If $\iota_m>r$ at the optimum, the constraint $s\geq\underline{s}(\iota_m,A)$ is binding, for otherwise one can improve welfare by decreasing both ι_m and s without changing y. A mentioned, if $\iota_m=r$ then $\underline{s}(\iota_m,A)=s(A)=\overline{s}(\iota_m,A)$ and the policy is inactive as either there is no excess reserve in equilibrium or $\iota=0$.

Theorem 4.1. *Assume (2) and that $v'(y)y$ is increasing. There exists a threshold $\tilde{A}>\bar{A}$ such that for all $A\leq\tilde{A}$, the optimal policy has $\pi>0$, $\iota>\pi$, and equilibrium $z>0$ larger than the upper bound in (22).*

Theorem 4.1 shows that whenever the economy hits the ZLB, it is optimal to intervene with a positive *real* IOER, and with a positive inflation rate to finance it. Under such intervention (monetary) equilibrium is unique. For A below $\bar{A}$, Lemma 3.1 (3) shows that equilibrium $y=y^*$ without intervention. Thus, informal trades are at the first-best levels and reducing them is marginally costless, while inflation can be used to finance returns on deposits through IOER which also reduce capital over-accumulation. Similarly, when $A>\bar{A}$ but close to it, it is relatively cheap to finance IOER and it is optimal to set ι_m above r. As a result, the optimal IOER is always positive. In contrast, when A is close to A^*, formal trades are close to their first-best levels with little capital over-accumulation, and setting $\iota_m=r$ is optimal. In fact, if the government is allowed to use fiscal instruments such as banking fees, it would be optimal to finance a deflationary policy.

We also remark that the result that inflation is optimal under ZLB will go through even if the central bank is endowed with some real assets whose proceeds can be used to finance IOER, as long as the IOER implementable without inflation is not too far from zero.

Theorem 4.1 shows the optimality of IOER with a flexible inflation target (that is set optimally). The following theorem reports the comparative statics for the optimal interest rate w.r.t. to productivity when a fixed inflation target is imposed.

Theorem 4.2. *Assume (2), that $v'(y)y$ is increasing, and that $\psi(k) = (k+\lambda)^x/(\lambda^{x-1}x) - \lambda/x$ for some $x > 1$ and $\lambda > 0$.*[14] *Suppose that the central bank is constrained to set $\pi = \bar{\pi} > 0$. Then, optimal $\iota > 0$ for all A and it increases with A. There exist four thresholds, $\bar{A}_{\bar{\iota}_m} < \hat{A} < \hat{\hat{A}} < A^*_{\bar{\iota}_m}$, where $\overline{s}(\bar{\iota}_m, A) = \bar{\iota}_m$ iff $A \leq \bar{A}_{\bar{\iota}_m}$, such that*

(1) for $A \geq A^*_{\bar{\iota}_m}$, the optimal $s = 0$;

(2) for $A \in (\hat{\hat{A}}, A^*_{\bar{\iota}_m})$, the optimal $s = \overline{s}(\bar{\iota}_m, A)$ and there is no equilibrium reserve holding;

(3) for $A < \hat{A}$, any optimal $s < \overline{s}(\bar{\iota}_m, A)$ and equilibrium reserve holding is positive.

Moreover, there exist $\bar{\sigma}_u > 0$ such that if $\sigma_u < \bar{\sigma}_u$, then any optimal $s < r$ (and hence optimal $\iota > \bar{\pi}$) for all A and optimal $s = \underline{s}(\bar{\iota}_m, A)$ for A small.

The thresholds $\bar{A}_{\bar{\iota}_m}$ and $A^*_{\bar{\iota}_m}$ correspond to $\bar{A}$ and A^* given by Proposition 3.1 under the targeted inflation $\bar{\pi}$: formal trades and capital accumulation are at their first-best levels for $A \geq A^*_{\bar{\iota}_m}$ while the economy exhibits ZLB with banks holding excess reserves for $A \leq \bar{A}_{\bar{\iota}_m}$ under $\bar{\pi}$ and $\iota = 0$. Theorem 4.2 then draws similar conclusions to Theorem 4.1: Theorem 4.2 (3) states that optimal $s < \overline{s}(\bar{\iota}_m, A)$ and hence optimal IOER is positive with equilibrium excess reserves larger than those without it when the economy is within or close to the ZLB region; Theorem 4.2 (2) states that, in contrast, for A close to $A^*_{\bar{\iota}_m}$, optimal ι is such that no banks hold excess reserves in equilibrium and hence the intervention is inactive.

Different from Theorem 4.1, however, even within the ZLB region, the optimal s can be higher than $\underline{s}(\bar{\iota}_m, A)$, as ι_m is not set optimally. Indeed, since a lower s would further decrease informal trades, for $\bar{\iota}_m$ large it is not optimal to set s as low as possible. This implies that it can be optimal to leave some of the seigniorage revenue to be distributed lump-sum even in the presence of excess reserves. Moreover, for $\bar{\iota}_m$ high or σ_u high, it can be the case that optimal $s > r$ and hence optimal $\iota < \bar{\pi}$. However, for σ_u sufficiently low, this can never happen, as shown in Figure 1. We remark that while all these results do not require the functional form assumed, the optimal police may not be unique and the comparative statics w.r.t. A requires the functional form to utilize arguments based on supermodularity. Since both $\overline{s}(\bar{\iota}_m, A)$ and $\underline{s}(\bar{\iota}_m, A)$ are decreasing in A, we only need to show that the objective

[14] Note that we ensure that $\psi'(0) = \psi_0'(0) + 1 = 1$ with this functional form.

function for the planner has decreasing differences over s and A. Nevertheless, multiplicity is rare and we can show that any selection of optimal ι is increases with A.

While the result that optimal ι increases as productivity rises seems in line with the current policy that central banks tend to increase interest rate as the economy recovers, the underlying rationale is very different from the conventional New Keynesian reasoning. In our model, ι increases optimally with A since the fundamental value of collateral increases with productivity and hence it is feasible for the government, everything else equal, to sustain a higher IOER.

This result also highlights a novel effects of IOER, the redistribution between deposit users and cash users. While interest on reserves is passed onto depositors, as some argue in policy debates,[15] the cash users are hurt by the inflation tax in absolute terms and are less favoured in relative terms. Indeed, the optimality of $s = \overline{s}(\overline{\iota}_m, A)$ when $A \geq \hat{\hat{A}}$ is due to this redistribution concern, since a lower s, while feasible and helping formal trades and capital over-accumulation, would further hurt informal trades that is already quite inefficient.

Although we have shown that a positive IOER is better than zero, this does not claim that such policy is optimal among all alternatives that are feasible and incentive compatible. We turn to optimal policy within a wide range of instruments next.

5 Optimal liquidity regulations

In our study of optimal IOER, the equilibrium capital holding per bank is always determined by $K(s, A)$, and, whenever $s > 0$, there is capital over-accumulation and the only channel to alleviate this inefficiency is through a lower s, or a higher IOER. However, both in theory and in practice, there are many other ways to decrease capital holdings, e.g., capital taxes or liquidity requirements. Here we consider all possible policies that are fiscally feasible and incentive compatible, and endogenously determine the optimal regulations/taxes.

In particular, we study the constrained-efficient allocation by considering the consolidated government budget constraint and allow for possible taxes, as long as they are incentive compatible and respect voluntary participation. Consumers are essentially anonymous and they can be taxed only through the banking system or through inflation tax (which already occurs in the last section). Banks that issue deposits have public records, but they can avoid taxation or regulations (such as liquidity requirement) if they do not issue deposits (and conduct their business with *private equity*).

[15]See Keister (2016)'s testimony before the US Congress that argues for this transmission mechanism as a defense of the policy.

In this environment, an allocation consists of

$$\langle q, y, (b, \varphi), \chi, k, z, \pi \rangle, \tag{28}$$

where q is the production per pairwise match with formal sellers, y is the underground production per underground match, b is the deposit issuance per bank, φ is the price for each deposit contract, χ is the price for money in terms of deposits in the DM money/deposit market, k is the capital holdings per bank, and z is the reserve holding per bank, and π is the rate of reserve/money creation (which can be positive or negative).

The mechanism design problem can be studied in two separate parts, one for the consumer bargaining (with formal and underground sellers), money holding, and deposit holding, and the other for banks' capital holding and deposit issuance. In terms of allocations, the first part proposes an allocation (y, q, m^h, b^h, χ) for consumers/sellers (where m^h and b^h denote consumer money and deposit holdings when leaving the CM), and the second (k, z, b) for banks. The two parts are connected by the spread s and the money/reserve creation rate π.

We begin with the consumer allocation. While we allow the consumer to use both deposit and money when meeting a formal seller, we only consider proposed trades that use deposits only in those meetings, and the proposed portfolio for an informal consumer after the DM money/deposit market consists of deposits alone. In Appendix B1 we show that these restrictions are with no loss of generality in terms of implementable allocation, and hence the constrained-efficient allocation is not affected either.[16] Moreover, given these restrictions, the proposed money and deposit holdings leaving the CM uniquely determine the portfolio of formal and informal consumers after the DM money/deposit market and *vice versa*. Below we describe proposed consumer allocations with portfolios for formal and informal consumers after the DM money/deposit market.

We follow Hu et al. (2009) and consider a simple game to implement trades in the DM. An outcome for the consumers consists of the portfolio of each formal consumer, b^f, the portfolio for each informal consumer, (b^i, m^i), both after the DM money/deposit market, and the respective trades, (q, d) and (y, p). Note that these also determine the portfolio leaving the CM, (b^h, m^h), which satisfies

$$b^h = (1 - \sigma_u)b^f + \sigma_u b^i, \; m^h = \sigma^u m^i. \tag{29}$$

For pairwise meetings with formal sellers, a mechanism is a mapping $o^d(b + m) = (q, d)$ that maps the consumer's sum of deposit and money holding to a trade (q, d) with $d \leq b + m$.

[16] We also consider banking fees levied on consumers there and show that it does not expand the implementable allocations.

Assuming that the agents' portfolios are all observable, the rule of the game is as follows.[17] First, both the consumer and the seller choose to accept the proposed trade according to the mechanism o^d. If both accept then the trade is executed; otherwise, the consumer makes a take-it-leave-it offer to the seller, a stage to ensure that the proposed trade is renegotiation proof, a pairwise core requirement. Similarly, for underground meetings, a mechanism is a mapping $o^m(m) = (y, p)$ that maps the consumer's money holding m to a trade (y, p). The rule of the game is analogous to that with formal sellers.

We say that the proposed consumer outcome, $[b^f, (b^i, m^i), y, q]$, is *implementable* if, for a given (s, π) and χ, it is optimal for the consumer to hold (m^h, b^h) leaving the CM (which are given by (29)) and for formal consumers to hold b^f and informal consumer (b^i, m^i), and to accept the proposed trades in the respective DM meetings.

Lemma 5.1. *For a given s and π, the outcome $[\chi, b^f, (b^i, m^i), y, q]$ is implementable for consumers if and only if $\chi = (1+\pi)/\varphi$ and the following conditions hold (where b^h, m^h are given by (29), $s = \varphi/\beta - 1$, $\iota_m = (1+\pi)/\beta - 1$):*

$$-sb^h - \iota_m m^h + \sigma[u(q) - d] + \sigma_u[v(y) - p] \geq 0, \tag{30}$$

$$-sb^i + \sigma[u(q) - d] \geq 0, \ \ v(y) - p - \frac{\iota_m - s}{1+s} m^i \geq 0, \tag{31}$$

$$b^f = b^h + \chi m^h = b^i + \chi m^i, \ \ b^f \geq d \geq q, \ \ m^i \geq p \geq y, \tag{32}$$

$$\begin{aligned} &q = q^* \ \textit{if}\ b^f \geq q^*, \ \ d = b^f \ \textit{otherwise}; \\ &y = y^* \ \textit{if}\ m^i \geq y^*, \ \ p = m^i \ \textit{otherwise}. \end{aligned} \tag{33}$$

The price χ is necessarily the same as the relative price in CM, $(1+\pi)/\varphi$. Since what really matters in the DM money/deposit market is the total wealth measured by $b + \chi m$, any deviation from that price will lead to zero demand for money or deposit. The first two terms in (30) express costs of holding deposits and money across periods, respectively, and the last two terms express trade surplus from formal meetings and trade surplus from informal ones for the consumer, respectively. Thus, if condition (30) does not hold, a consumer can simply leave the CM without any asset and have no DM trade to obtain higher payoffs (than the proposed allocation). The two conditions in (31) are used to deter two other deviating portfolios a consumer can hold at different timings. On the one hand, if the first one does not hold, a consumer would prefer to leave the CM with real balances m^i and trade with informal sellers only. On the other hand, if the second inequality does not hold, an informal consumer

[17] The result does not change if we assume that agents are allowed to hide their asset holdings. Note also that although we only consider proposed outcomes where formal consumers hold deposits only, we need to consider deviations to all possible portfolios (including those containing money holding) and show that the proposed holdings are optimal for consumers.

would deviate to sell all his money for deposits and skip the meeting. Our proof of sufficiency shows that these are the only relevant deviations. The first condition in (32) follows from consumer budget constraint, while the rest follows from their liquidity constraint (that they need to use either deposit or money to finance their consumption) and seller (formal and informal) participation constraints. Finally, (33) ensures that pairwise core is satisfied in both meetings with formal and informal sellers.

For banks an outcome is a triple (k, z, b), where b is each bank's deposit issuance, k capital holding, and z reserve holding. Both z and b are expressed in terms of coming CM goods. Here we consider two types of policies for implementation; in the Appendix B1 we show that this is without loss of generality as long as voluntary participation is required.

The first policy has no explicit taxation but uses inflation, while the second uses explicit taxation. Under the first policy, we have $\pi > 0$, and the other policy instruments include (a) a fixed liquidity requirement, $\bar{z}$, that a bank has to hold regardless of its balance sheet;[18] (b) a proportional liquidity requirement so that a bank needs to hold ηk reserves when holding k units of capital; and (c) interest on excess reserves, ι. Both liquidity requirements are in terms of next-period CM goods, and hence, to hold z units of reserves, a bank needs to pay

$$\frac{1+\pi}{1+\iota}(z - \bar{z} - \eta k) + (1+\pi)(\bar{z} + \eta k)$$

units of CM goods today. There is also a budget constraint analogous to (24), but taking the required reserves into account. Under the second policy, π can be positive or negative, and the policy instruments include (a) a fixed fee on banking, τ_0; (b) a proportional tax on capital holding with rate τ_1; and (c) interest on excess reserves, ι. Both taxes are in terms of current-period CM goods. We allow for negative τ_0 and τ_1, interpreted as subsidies, but, as will be seen later, subsidies are never used for constrained-efficient allocations.

An outcome (k, z, b) is *implementable* under parameters (s, π, m^h) if it is the optimal choice for a bank when the competitive price for bank deposit is $\varphi = \beta(1+s)$, the prevailing inflation rate (for both money and reserve) is π, and money holding is m^h per consumer, subject to the following participation constraint. Banks can opt out by not issuing any deposits. In doing so, they can avoid all the regulation mentioned above. In that case, their profit will be given by

$$\Pi_A^* \equiv \beta\{Ak^*(A) - (1+r)\psi[k^*(A)]\} = \psi'[k^*(A)]k^*(A) - \psi[k^*(A)]. \tag{34}$$

[18]This fixed liquidity requirement is similar to the scheme proposed by Andolfatto (2010) and Wallace (2014) in a pure-currency economy. Their scheme pays interest on money to agents, but only if the agent's money holding exceeds certain amount (which would be analogous to liquidity requirement here). While our scheme shares similar economic intuition, ours is implemented through the banking system.

Note that, by convexity of ψ, Π_A^* increases with A since $k^*(A)$ does.

Lemma 5.2. *For a given s, π and m^h, an outcome (k, z, b) that satisfies $b = \rho Ak + z$ is implementable if and only if*

$$sb - rz + Ak - (1+r)\psi(k) + \frac{(1+r)\pi m^h}{n} \geq (1+r)\Pi_A^*. \tag{35}$$

The first term in the left-side of (35) is the profit from issuing deposits, and the second is the cost of holding reserves across period. The third and the fourth are the benefit and the cost of holding capital. The last term comes from the fact that the government uses seigniorage revenue to finance the interest on reserves, and it includes such revenue from consumers' money holdings.[19]

Finally, market clearing and pledgeability require

$$b^h = nb \leq n(\rho Ak + z).$$

We can then summarize conditions for implementability as follows.

Theorem 5.1. *An allocation, $\langle q, k, y, z, m^h \rangle$, is implementable only if the following hold:*

$$\begin{aligned} -r(nz + m^h) + n[Ak - (1+r)\psi(k)] + \sigma_u[v(y) - y] + \sigma[u(q) - q] &\geq n(1+r)\Pi_A^*, && (36)\\ n\rho Ak + (nz + m^h) + \sigma_u[v(y) - y] &\geq q, && (37)\\ m^h \geq \sigma_u y,\ z &\geq 0. && (38) \end{aligned}$$

If, in addition to the above conditions, (a) $m^h = \sigma_u y$, (b) $q \geq v(y)$, and (c) (37) holds with equality or $q = q^*$, then it is implementable.

Note that the condition (36) is free from the policy variables, and is derived from summing up (30) and (35), together with the CM market clearing condition, by appropriate coefficients to cancel out s and ι_m. Condition (37) is directly derived from the second condition in (31), (32), and the CM market clearing condition. As such, it is necessary for implementation, and either banks or consumers would have profitable deviations unless these conditions are met. The sufficiency is proved by constructing the appropriate policy variables from the two schemes mentioned above. The additional condition (a) for sufficiency is to ensure the consumers have sufficient deposits to withdraw for trades with informal sellers, while conditions (b) and (c) are to ensure the pairwise core requirement. Now, if $-rq^* + \sigma[u(q^*) -$

[19]Since banks have public records, to increase their liquidity provision, one can allow them to issue unsecured deposits disciplined by their future profits, or the *franchise value*. However, as we show in Appendix B1, this does not expand implementable allocations.

$q^*] \geq 0$, then the first-best level of trades in formal meetings can be implemented without banks but with fiat money, and this would be an uninteresting case. Thus, we assume

$$r > r_0 \equiv \frac{\sigma[u(q^*) - q^*]}{q^*}. \tag{39}$$

Using Theorem 5.1, we find the constrained-efficient allocation as follows. First we maximize the welfare (3) subject to the constraints (36)-(38). Note that since z and m^h appear in (36) and (37) through $nz + m^h$, it is with no loss of generality to assume that $m^h = \sigma_u y$ and reduce (38) to $z \geq 0$. The solution is unique as this is a convex problem, the solution denoted by (q^e, k^e, y^e, z^e). Since the constraints (36)-(38) are necessary for implementation, we only need to show that (q^e, k^e, y^e, z^e) is in fact implementable. We do so by demonstrating that the additional conditions (a)-(c) in Theorem 5.1 are satisfied by (q^e, k^e, y^e), and we need the assumptions (2) and (39) for this step. The following theorem characterizes the constrained-efficient allocation and the optimal policy to implement it.

Theorem 5.2. *Assume (2) and (39). There is a unique constrained efficient allocation, denoted by (q^e, k^e, y^e, z^e). There exist thresholds $A^{**} < A^*$ and $\hat{A}$ such that*

(1) if $A \geq A^{**}$, $(q^e, k^e, y^e) = (q^*, k^*(A), y^*)$; otherwise, $q^e < q^*$, $k^e > k^*$, and $y^e \leq y^*$.

(2) if $A > \hat{A}$, optimal $\pi \leq 0$ and, whenever the first-best is not implementable, optimal $\tau_0 > 0$ and $\tau_1 > 0$;

(3) if $A < \hat{A}$, optimal $\pi > 0$ and the constrained efficient allocation can be implemented without explicit taxation, and, whenever the first-best is not implementable, optimal $\eta > 0$ and $\bar{z} > 0$; the optimal s and the optimal inflation both strictly decrease with A.

Finally, $\hat{A} > A^{**}$ if and only if

$$\frac{v(y^*)}{y^*} \frac{\{\sigma[u(q^*) - q^*] + q^*\}}{q^*} > 1 + r. \tag{40}$$

According to Theorem 5.2 (1), under the optimal trading mechanism and intervention, the first-best allocation is implementable when the productivity is high (which allows for abundant collateral). Even in that case, however, intervention may be necessary, especially so when $v(y^*) > (1 + r)y^*$. In that case, deflation is needed when A is high. Reserves may not be essential, however, as the optimal scheme uses proportional taxes on capital to finance deflation while there is already plenty of capital to back sufficient deposit issuance for q^*. In contrast, when $A < A^{**}$, the constrained efficient allocation features $q^e < q^*$ and $k^e > k^*(A)$. Hence, it is optimal to have capital over-accumulation when liquidity in formal trades is constrained. Moreover, we can show that for small σ_u, the optimal intervention always requires banks to hold excess reserves and to pay interest on them. Thus, nominal assets are essential for welfare. This gives a novel perspective on coexistence: even at the limit case where σ_u goes to zero, money in the form of reserves is still essential for welfare.[20]

Theorem 5.2 (2) and (3) give a full characterization of the optimal intervention to implement the constrained-efficient allocation. Similar to Theorem 4.1, there is always a range of productivity levels, which happen to be on the lower end, under which inflation is optimal and the optimal intervention uses IOER. Different from that theorem, however, with optimal trading mechanism and optimal intervention, inflation can be optimal even when the first-best is implementable and hence it widens the range of situations for inflation to be optimal. Moreover, it shows that the new tools identified here, the proportional and fixed liquidity requirement without interest payments, are essential for welfare. Whenever inflation is optimal, those tools are active to implement the constrained-efficient allocation, except for the extremely high discount factor under which informal trades can achieve first-best without any intervention (that is, when $r \leq r_1 \equiv v(y^*)/y^* - 1$).

Theorem 5.2 also shows that intervention is always optimal, unless the first-best is implementable. In fact, as mentioned, we can use free market only if $r \leq r_1$. Otherwise, either liquidity management is needed or explicit taxation is called for. This is remarkable as we have already adopted the optimal trading mechanism, and hence these results demonstrated that the intervention is used to tackle the underlying frictions such as limited commitment.

6 Conclusion and Extensions

We developed a model with endogenous liquidity needs which are provided by both privately issued deposit contracts with pledged capital holdings and government issued money and reserves. Without government intervention, when the productivity is low and hence pledgeable assets are scarce, banks hold excess reserves in equilibrium and nominal interest rate becomes zero. In this circumstance, we demonstrated the optimality of IOER financed by inflation taxes. Moreover, we showed that ZLB would not occur under optimal policy. We also considered all fiscally feasible and incentive compatible policies that respect voluntary participation and appropriate core requirements, and showed that, in addition to IOER, proportional and fixed liquidity requirements are optimal when productivity is not too high. These requirements can be interpreted as requirements on liquidity coverage ratios, which

[20]This result is similar to Hu and Rocheteau (2013), who point out that it is optimal to incentivize agents to hold money when capital can be used as means-of-payments. As here, the main inefficiency money corrects is capital over-accumulation.

are key regulations in the Basel III. These results demonstrate the optimality of the IOER policy and suggest a holistic approach to both monetary policy and banking regulations to achieve constrained-optimal allocations. Since these interventions are optimal even under the optimal trading mechanism, they are *essential* for the welfare in the sense that they are not driven by suboptimal trading mechanisms but derived from the underlying frictions such as lack of monitoring and limited commitment. Finally, our results also suggest that the optimal real interest rate on reserves is procyclical, and this may provide some guidance on the future of monetary policy based on interest on reserves.

Below we discuss three extensions of our model. The first considers negative externality of informal trades, the second narrow banking, and the third transition dynamics.

Negative externality of underground trades

Some economists have argued that the secrecy of currency trades allow for many undesired transactions that would otherwise be infeasible (see, for example, Sands (2016) and Rogoff (2016)). One simple way to incorporate this in our model is to assume that the underground consumption causes some negative externality to other consumers, and this changes the term (c) in the welfare expression (3) into

$$\sigma_u[v(y) - y] - \mathcal{E}\sigma_u y, \tag{41}$$

where $\mathcal{E}\sigma_u y$ measures the negative externality occurred to all consumers when the average consumption from underground meetings per consumer in $\sigma_u y$.[21] Note that this extension does not affect any equilibrium analysis, nor implementable allocations, but only changes the social welfare function.

Our analysis in the main text is then a special case with $\mathcal{E} = 0$. We describe below how our results will change when $\mathcal{E}$ becomes positive. When we restrict the policy instruments to IOER and reserve/money creation, as we did in Section 4, a positive $\mathcal{E}$ implies a higher $\tilde{A}$ in Theorem 4.1. In particular, if $\mathcal{E} \geq r$, then it is always optimal to use inflation to finance a positive IOER, regardless of the level of productivity, A. Under optimal trading mechanism and intervention, a positive $\mathcal{E}$ would mostly alter the constrained efficient y^e; indeed, in this case at the first best $\tilde{y}$ we have $v'(\tilde{y}) = 1 + \mathcal{E} > 1$ and hence $\tilde{y} < y^*$. It can be shown that the constrained efficient allocation features $y^e \in [\tilde{y}, y^*]$, and the threshold $\hat{A}$ in Theorem 5.2 would increase with $\mathcal{E}$. In particular, when $\mathcal{E} \geq r$, $\hat{A} = \infty$ for any utility function. Moreover,

[21] Of course, in reality some of the cash trades are completely legitimate while others are socially undesirable. However, note that since there is no record keeping or monitoring in these trades one cannot distinguish the legitimate use from others. Moreover, one could introduce heterogeneous underground meetings but that would not change any substance.

in this case the optimal nominal IOER increases with A. See Appendix B2 for details.

Narrow banking

Since IOER is implemented, many (see for example, Cochrane (2014)) have advocated for the use of narrow banks, ones that hold central bank reserves and issue deposits, or a narrow banking system, where all depository institutions hold only very safe assets such as government liabilities. We discuss these proposals in our framework here. First, we can allow consumers to hold reserves and receive interest on reserves. However, this would not be in general optimal. One crucial element in our optimal intervention design is to use reserve requirements to implicitly tax banks so that IOER can be financed more efficiently, but allowing for narrow banks would interrupt this intervention without expanding implementable allocations. In turn, a narrow-banking system in our framework would entail a policy that prohibits capital-holding firms from issuing deposits. Instead, all depository institutions can only hold nominal assets. Since it is always feasible to induce $k = k^*(A)$ with the full set of feasible policy instruments but the constrained efficient allocation features $k > k^*(A)$ unless the first-best is implementable, narrow banking is suboptimal.

Nevertheless, narrow banking could be potentially optimal if not all incentive compatible and feasible interventions are available due to, for example, political economy reasons. For example, suppose that IOER can only be financed through inflation. When $\sigma_u = 0$, under narrow banking, the DM allocation is always given by $q = Q(r)$, while the equilibrium capital holding is $k = k^*(A)$ per firm. This is a better allocation than under private liquidity provision (as characterized by Lemma 3.1 (3)) when $A < \bar{A}$ and without any intervention, as private liquidity provision induces capital over-accumulation. By continuity, there exists some $\tilde{A} \in (\bar{A}, A^*)$ such that narrow banking delivers a higher welfare than private liquidity provision by firms if $A < \tilde{A}$. This welfare raking does not change for small σ_u's and IOER allowed, as inflation tax available depends continuously on σ_u. That is, and narrow banking can be better if σ_u is small and A is low, when interventions are limited due to reasons other than incentive-compatibility.

Transition dynamics

We conducted comparative statics for different stationary equilibria under different parameters (such as capital productivity, A). The qualitative results will remain, such as the optimality of IOER and inflationary policies, if we consider a Markovian environment where A is persistent. Moreover, although our model features quasi-linearity in capital production and hence the banks can fully readjust their capital holdings, the results would not change if capital accumulation is constrained by capital output, as in Venkateswaran and Wright (2013). However, when allowing for shocks to productivity and limited capital accumulation, the model can be used to study its implications for inflation dynamics and investment dynamics, and their interactions with the policies.

In particular, since our model predicts that excess reserve holdings would increase as productivity drop, in the Markovian environment where productivity varies over time the demand for nominal assets would increase when productivity is low while it would decrease when productivity is high. This suggests that inflation will be procyclical, a Phillips curve type of result. In our analysis, which assumes a stationary environment, the optimal long-run inflation rate does not take this transitory implications for the price levels into account. A full analysis of these dynamics and their interactions with optimal intervention, however, is beyond the scope of this paper.

References

[1] Andolfatto, D. (2010). "Essential Interest-bearing Money," *Journal of Economic Theory* 145, 1495-1507.

[2] Cochrane, John (2014). "Toward a Run-free Fnancial System,"in *Across the Great Divide: New Perspectives on the Financial Crisis*, edited by Martin Neil Baily and John B. Taylor, Hoover Institution Press, 197-249.

[3] Goodfriend, Marvin (2002). "Interest on Reserves and Monetary Policy," *FRB Economic Policy Review* vol 8.

[4] Hu, Tai-Wei, John Kennan, and Neil Wallacee (2009). "Coalition-Proof Trade and the Friedman Rule in the Lagos-Wright Model," *Journal of Political Economy* 117, 116-137.

[5] Hu, Tai-Wei and Guillaume Rocheteau (2013). "On the Coexistence of Money and Higher-return Assets and its Social Role," *Journal of Economic Theory* 148, 2520-2560.

[6] Keister, Todd (2016). "Interest on Reserves," Transcript of Oral Testimony delivered before the Subcommittee on Monetary Policy and Trade Committee on Financial Services, United States House of Representatives.

[7] Kiyotaki, Nobuhiro and John Moore (1997). "Credit Cycle," *Journal of Political Economy* 105, 211-248.

[8] Lagos, Ricardo and Randall Wright (2005). "A Unified Framework for Monetary Theory and Policy Analysis," *Journal of Political Economy* 113, 463-484.

[9] Lagos, Ricardo and Guillaume Rocheteau (2008). “Money and Capital as Competing Media of Exchange,” *Journal of Economic Theory* 142, 247-258.

[10] Rachel, Lukasz and Lawrence Summers (2019). “On Falling Natural Rates, Fiscal Policy, and the Risk of Secular Stagnation,” *Brookings Papers on Economic Activity*.

[11] Rocheteau, Guillaume and Randall Wright (2005). “Money in Search Equilibrium, in Competitive Equilibrium, and in Competitive Search Equilibrium,” *Econometrica* 73, 175-202.

[12] Rogoff, Kenneth S. (2016). *The Curse of Cash*, Princeton University Press.

[13] Sands, Peter (2016). “Making it Harder for the Bad Guys: The Case for Eliminating High Denomination Notes ”*M-RCBG Associate Working Paper Series, No. 52*.

[14] Stein, Jeremy C. (2012). “Monetary Policy as Financial Stability Regulation,” *Quarterly Journal of Economics*, 127, 57-95.

[15] Topkis, M. Donald (1998). *Supermodularity and Complementarity,* Princeton University Press.

[16] Venkateswaran, Vaidyanathan and Randall Wright (2013). “Pledgeability and Liquidity: A New Monetarist Model of Macro and Financial Activity, ”*NBER Macroeconomics Annual,* 28.

[17] Wallace, Neil (2001). “Whither Monetary Economics?”*International Economic Review*, 42, 847-869.

[18] Wallace, Neil (2014). “Optimal Money-creation in ‘Pure-Currency’ Economies: a Conjecture.” *Quarterly Journal of Economics*, 129, 259-274.

[19] Williamson, Stephen D. (2012). “Liquidity, Monetary Policy, and the Financial Crisis: A New Monetarist Approach,” *American Economic Review* 102, 2570-2605.

Appendix A: Proofs

We begin our proofs with a few facts that will be referred to later.

(F1) The function $Q(s)$ is strictly decreasing in s with $Q(0) = q^*$, and $Y(\iota_m, s)$ is (strictly) increasing in s but (strictly) decreasing in ι_m.

(F2) The function $K(s, A)$ is strictly increasing in both s and A, with $K(0, A) = k^*(A)$.

(F3) Assume that $v'(y)y$ is (strictly) increasing. The function $\frac{1+\iota_m}{1+s} Y(\iota_m, s)$ is (strictly) increasing in s but (strictly) decreasing in ι_m.

(F1) and (F2) follow directly from the strict concavity of u and v, and strict convexity of ψ. For (F3), since $v'[Y(\iota_m, s)] = (1 + \iota_m)/(1 + s)$,

$$\frac{1 + \iota_m}{1 + s} Y(\iota_m, s) = v'[Y(\iota_m, s)]Y(\iota_m, s),$$

which is strictly increasing in $Y(\iota_m, s)$. The result then follows from (F1).

Proof of Lemma 3.1

Proof. Since $s \in [0, r)$, $\chi = 1/\varphi \in (1, 1/\beta]$. We first verify that the portfolio $b = D(s)$ given by (18) and $m = M(s)$ is optimal for CM and satisfies the market clearing for the DM money/deposit market, (15). Note that the first condition in (15) is satisfied by $m^i = Y(s)$ and $m^f = 0$ with $m = M(s)$, and the second is automatically satisfied by Walras' Law. For optimality, by (9), to show that in equilibrium $v'(y) = \chi$, we only need to verify that $Y(s) \leq (1/\chi)D(s) + M(s)$, and , by (18), this is equivalent to (note that $\chi = 1/\varphi = (1+r)/(1+s)$)

$$Q(s) \geq \frac{1 + r}{1 + s} Y(s). \tag{42}$$

Since $v'[Y(s)] = \chi = 1/\varphi = (1+r)/(1+s)$, this is further equivalent to $Q(s) \geq v'[Y(s)]Y(s)$. By (F1) and (F3) (note that $Y(s) = Y(r, s)$), it suffices to show that

$$Q(r) \geq v'[Y(r)]Y(r) = v'(y^*)y^* = y^*. \tag{43}$$

By definition of $Q(s)$ and by (2),

$$u'[Q(r)] \quad = \quad \frac{\sigma + r}{\sigma} < 1/\theta = u'(y^*),$$

and hence $Q(r) > y^*$ by concavity of u.

Now we consider the case where $s = r$. Since $\phi = 1 = \chi$ in this case, formal consumers are indifferent between deposits and money in the DM money/deposit market, but informal consumers need at least $Y(r)$ to leave that market to finance informal trades. The earlier argument that $v'(y) = \chi$ is still valid as long as (2) holds, and the demand for money is at least $M(r) = \sigma_u Y(r)$ with the optimal informal consumption equal to $Y(r) = y^*$. Similarly, optimal formal consumption is $Q(r)$. However, the consumer can hold more money than $M(r)$ in the CM and less deposits, as long as the total value is equal to $Q(r)$. $\square$

Proof of Proposition 3.1

Proof. (1) Since $A \geq A^*$ and since $k^*(A)$ is increasing in A, by (20),

$$Q(0) = q^* \leq n\rho A k^*(A) + \sigma_u(1 + r)Y(0), \tag{44}$$

and hence (19) is satisfied with $s = 0$ and $z = 0$.

(2) Now, $A < A^*$ implies that (44) fails. However, since the left-side of (19) is strictly decreasing in s with $Q(\infty) = 0$ but the right-side is strictly increasing with $z = 0$, there is a unique $s(A) > 0$ such that (19) holds with equality. By (F3) $Y(s)/(1 + s)$ is increasing in s. $A \geq \bar{A}$ ensures that $s(A) \in (0, r]$. Finally, since $K(s, A)$ is strictly increasing in A, the right-side of (19) strictly increases with A, and hence $s(A)$ strictly decreases with A (see also the left panel of Figure 2).

(3) Now consider $A < \bar{A}$. First we show that equilibrium $s = r$. Such equilibrium exists since at $s = r$, banks are willing to hold any amount of reserves. Let

$$nz = Q(r) - n\rho A K(r, A) - \sigma_u Y(r) > 0, \tag{45}$$

where the inequality follows from $A < \bar{A}$. Then each bank holds z is an equilibrium. Also, in equilibrium s must be equal to r: on the one hand, $s > r$ cannot be an equilibrium, for otherwise there will be infinite demand for reserves from banks; on the other hand, if $s < r$, the demands for deposits and for money are uniquely pinned down, and no bank is willing to hold reserves, but $A < \bar{A}$ implies that (19) cannot be satisfied.

As argued in Lemma 3.1, the minimum demand for real balances is $M(r) = \sigma_u Y(r)$, and the formal consumption is $Q(r)$. Then, z determined by (45) is the largest equilibrium real value of reserves. However, any real balance holding between $M(r)$ and $M(r) + nz$ is also an equilibrium (with corresponding lower reserves in real terms). Since $K(s, A)$ is strictly increasing in A, z determined by (45) is strictly decreasing in A (see also the right panel in Figure 2). $\square$

Proof of Lemma 4.1

Proof. Let ι, ι_m, and s satisfying (26) be given. Following the same argument as in Lemma 3.1, equilibrium money holding per consumer will be $\sigma_u Y(\iota_m, s)$. By (24), equilibrium reserve holding (in per consumer terms) has to satisfy

$$\hat{z} \leq \frac{\pi(1+\iota)}{\iota - \pi}\sigma_u Y(\iota_m, s) = \frac{\iota_m - r}{r - s}\sigma_u Y(\iota_m, s), \tag{46}$$

where the equality follows from (26). By market clearing,

$$Q(s) \leq n\rho A k + \hat{z} + \frac{1+\iota_m}{1+s}\sigma_u Y(\iota_m, s),$$

the result follows immediate from the banks' profit maximization and (46). □

Proof of Theorem 4.1

Proof. Let $A < A^*$. If at the optimum $\iota_m > r$, then $s = \underline{s}(\iota_m, A)$. Indeed, if $\iota_m > r$ and $s > \underline{s}(\iota_m, A)$, then one can decrease both ι_m and s so that $(1+\iota_m)/(1+s)$ is kept at constant, and hence $Y(\iota_m, s)$ is kept at constant, but a lower s leads to higher $q = Q(s)$ and lower $k = K(s, A)$, and this improves welfare.

Our strategy is then to exploit this constraint and reduce the policy variable to ι_m alone, and prove the result by computing the derivative w.r.t. ι_m. However, we need to be careful with the bounds on ι_m. For the lower bound, note that $\overline{s}(r, A) = s(A)$ given by Proposition 3.1 for any $A < A^*$. Now we show that, for any $A < A^*$,

$$\begin{aligned} &\lim_{\iota_m \downarrow r} \underline{s}(\iota_m, A) = s(A), \\ &\lim_{\iota_m \downarrow r} \sigma_u \frac{(\iota_m - r)Y[\iota_m, \underline{s}(\iota_m, A)]}{r - \underline{s}(\iota_m, A)} = Q[s(A)] - \sigma_u Y[r, s(A)] - n\rho A K[s(A), A]. \end{aligned} \tag{47}$$

In particular, for $A < \bar{A}$, the latter converges to the upper bound of nz given by Proposition 3.1 (3). The convergence is straightforward for $A > \bar{A}$ as the term $r - \underline{s}(\iota_m, A)$ converges to $r - s(A) > 0$. For $A \leq \bar{A}$, recall that $s(A) = r$, and suppose, by contradiction, that there is decreasing a sequence $\{\iota_m^\nu\}_{\nu=0}^\infty$ with limit r such that $\lim_{\nu\to\infty} \underline{s}(\iota_m^\nu, A) = \hat{s} < r$. Then,

$$\begin{aligned} &\lim_{\nu\to\infty} -Q(s^\nu) + n\rho A K(s^\nu, A) + \sigma_u \left\{ \frac{\iota_m^\nu - r}{r - s^\nu} + \frac{1+\iota_m^\nu}{1+s^\nu} \right\} Y(\iota_m^\nu, s^\nu) \\ = \;& -Q(\hat{s}) + n\rho A K(\hat{s}, A) + \sigma_u \frac{1+r}{1+\hat{s}} Y(r, \hat{s}) < -Q(r) + n\rho A K(r, A) + \sigma_u \frac{1+r}{1+r} Y(r, r) \leq 0. \end{aligned}$$

where $s^\nu = \underline{s}(\iota_m^\nu, A)$ and the equality follows from continuity, the first inequality from $\hat{s} < r$ and (F1)-(F3), and the second from $A \le \bar{A}$. This is a contradiction to the definition of $\underline{s}$, under which the limit should be zero as each element in the sequence is zero. Since any such sequence $\{s^\nu\}$ is bounded from above by r, this implies (47). Thus, we may define $\underline{s}(r, A) = r$ for $A \le \bar{A}$, which would preserve continuity.

Thus, we can reduce the policy variable to $\iota_m \in [r, \bar{\iota}_m]$ and set $s = \underline{s}(\iota_m, A)$, and the central bank's problem is to maximize

$$\begin{aligned} G(\iota_m; A) &= \sigma\{u[Q(\underline{s}(\iota_m, A))] - Q(\underline{s}(\iota_m, A))\} + n\{AK[\underline{s}(\iota_m, A), A] - (1+r)\psi[K(\underline{s}(\iota_m, A), A)]\} \\ &+ \sigma_u\{v[Y(\iota_m, \underline{s}(\iota_m, A))] - Y(\iota_m, \underline{s}(\iota_m, A))\}. \end{aligned} \tag{48}$$

We show that there exists $\tilde{A} > \bar{A}$ such that for all $A \le \tilde{A}$, $\frac{\partial}{\partial \iota_m} G(r; A) > 0$. This implies that optimal $\iota_m > r$ for all $A \le \tilde{A}$. Since it is optimal to set $s = \underline{s}(\iota_m, A)$, this implies that equilibrium reserves is larger than the upper bound given by (22) for $A < \bar{A}$ and strictly positive for $A \in [\bar{A}, \tilde{A}]$.

Now, we can compute the derivative $\frac{\partial}{\partial \iota_m} G(\iota_m; A)$ as

$$\begin{aligned} &\frac{\partial}{\partial \iota_m} G(\iota_m; A) = \{\sigma[u'(Q) - 1]Q_s + n[A - (1+r)\psi'(K)]K_s\}\, \underline{s}_{\iota_m} + \sigma_u[v'(Y) - 1](Y_{\iota_m} + Y_s \cdot \underline{s}_{\iota_m}) \\ = \quad &\left\{ \{\sigma[u'(Q) - 1]Q_s + n[A - (1+r)\psi'(K)]K_s\} + \sigma_u[v'(Y) - 1]\left(\frac{Y_{\iota_m}}{\underline{s}_{\iota_m}} + Y_s \right) \right\} \underline{s}_{\iota_m}, \end{aligned} \tag{49}$$

where $s = \underline{s}(\iota_m, A)$, $\underline{s}_{\iota_m} = \frac{\partial}{\partial \iota_m}\underline{s}$, $Y_s = \frac{\partial}{\partial s} Y$, $Y_{\iota_m} = \frac{\partial}{\partial \iota_m} Y$, $Q_s = \frac{\partial}{\partial s} Q$, and $K_s = \frac{\partial}{\partial s} K$. To sign the derivative at $\iota_m = r$, we first compute $\frac{\partial}{\partial \iota_m}\underline{s}(\iota_m, A)$ by Implicit Function Theorem:

$$\frac{\partial}{\partial \iota_m}\underline{s}(\iota_m, A) = \frac{-\sigma_u\{Y + (\iota_m - r)Y_{\iota_m} + (r-s)[v''(Y)Y + v'(Y)]Y_{\iota_m}\}}{(r-s)(-Q'(s) + n\rho A K_s) + \sigma_u\left\{\frac{\iota_m - r}{r-s} Y + (\iota_m - r)Y_s + (r-s)[v''(Y)Y + v'(Y)]Y_s\right\}}$$

for all $\iota_m > r$; we show that $\frac{\partial}{\partial \iota_m}\underline{s}(\iota_m, A) < 0$ for ι_m close to r and is bounded away from zero at the limit. Now, by (47), taking ι_m to r from above,

$$\lim_{\iota_m \downarrow r} \frac{\partial}{\partial \iota_m}\underline{s}(\iota_m, A) = \begin{cases} \frac{-\{Y + [r - s(A)][v''(Y)Y + v'(Y)]Y_{\iota_m}\}}{[r - s(A)]\{-Q_s + n\rho K_s + \sigma_u[v''(Y)Y + v'(Y)]Y_s\}} \text{ if } A > \bar{A}, \\ \lim_{\iota_m \downarrow r} \frac{-\sigma_u Y}{\sigma_u \frac{\iota_m - r}{r - \underline{s}(\iota_m, A)} Y(r,r)} = \frac{-\sigma_u Y(r,r)}{Q(r) - \sigma_u Y(r,r) - n\rho A K(r, A)} \text{ if } A \le \bar{A}, \end{cases}$$

where the derivatives are evaluated at $s = s(A)$ (note that $s(A) = r$ for $A \le \bar{A}$). Clearly, $\frac{\partial}{\partial \iota_m}\underline{s}(r, A) < 0$ when $A \le \bar{A}$; in fact, $\frac{\partial}{\partial \iota_m}\underline{s}(r, \bar{A}) = -\infty$. For $A > \bar{A}$, the denominator is positive by (F1)-(F3), and the numerator is negative if $Y + [r - s(A)][v''(Y)Y + v'(Y)]Y_{\iota_m} > 0$, which holds strictly at $A = \bar{A}$ as $s(\bar{A}) = r$ and hence, by continuity, it holds for A close to $\bar{A}$

but lightly larger, say below the threshold $\tilde{A}_1$. This shows that $\frac{\partial}{\partial \iota_m}\underline{s}(\iota_m, A) < 0$ for $A \leq \tilde{A}_1$.

Finally, moving back to (49) evaluated at $\iota_m = r$ (recall that then $\underline{s}(r, A) = s(A)$), the term $\{\sigma[u'(Q) - 1]Q_s + n[A - (1 + r)\psi'(K)]K_s\}$ is strictly negative for $A < A^*$ and $s = s(A)$ by (F1)-(F2) and the fact that $K(s, A) > k^*(A)$. For $A \leq \bar{A}$, the term $\sigma_u[v'(Y) - 1] = 0$ since $\iota_m = r$, and since $\frac{\partial}{\partial \iota_m}\underline{s}(r, \bar{A}) < 0$, it follows that $\frac{\partial}{\partial \iota_m}G(r; A) > 0$. Moreover, by continuity, for A close to $\bar{A}$ but slightly larger, we have

$$\{\sigma[u'(Q) - 1]Q_s + n[A - (1 + r)\psi'(K)]K_s\} + \sigma_u[v'(Y) - 1]\left(\frac{Y_{\iota_m}}{\underline{s}_{\iota_m}} + Y_s\right) < 0$$

(note that $\underline{s}_{\iota_m}$ is bounded away from zero), say below the threshold $\tilde{A}_2$. Take $\tilde{A} = \min\{\tilde{A}_1, \tilde{A}_2\}$. Then, $\frac{\partial}{\partial \iota_m}G(r; A) > 0$ for all $A \leq \tilde{A}$. □

Proof of Theorem 4.2

Proof. Let $\bar{\iota}_m = (1 + r)(1 + \bar{\pi}) - 1$. The proof first deals with Theorem 4.2 (1)-(3), then shows that for σ_u small, optimal $s < r$ for all A and optimal $s = \underline{s}(\bar{\iota}_m, A)$ for all A small, and, finally, shows that optimal ι increases with A.

Part 1 Define $A^*_{\bar{\iota}_m}$ and $\bar{A}_{\bar{\iota}_m}$ as the unique solution to

$$\begin{aligned} Q(0) &= \rho A^*_{\bar{\iota}_m} k^*(A^*_{\bar{\iota}_m}) + (1 + \bar{\iota}_m)Y(\bar{\iota}_m, 0) \\ \text{and } Q(\bar{\iota}_m) &= \rho \bar{A}_{\bar{\iota}_m} K(\bar{\iota}_m, \bar{A}_{\bar{\iota}_m}) + Y(\bar{\iota}_m, \bar{\iota}_m), \end{aligned}$$

respectively. For $A \geq A^*_{\bar{\iota}_m}$, only $s = 0$ is implementable and the optimal $\iota = r$ with no equilibrium excess reserve holdings. This proves Theorem 4.2 (1).

Part 2 Now consider $A < A^*(\bar{\iota}_m)$, and the central bank chooses s to maximize

$$G(s, A) \equiv \sigma\{u[Q(s)] - Q(s)\} + \sigma_u\{v[Y(\bar{\iota}_m, s)] - Y(\bar{\iota}_m, s)\} + n\{AK(s, A) - (1 + r)\psi[K(s, A)]\}$$

subject to (27), that is, $s \in O(A) \equiv [\underline{s}(\bar{\iota}_m, A), \bar{s}(\bar{\iota}_m, A)]$. We prove Theorem 4.2 (2) and (3) by showing the following:

(a) There exists $\hat{\hat{A}} < A^*_{\bar{\iota}_m}$ such that for all $A \in [\hat{\hat{A}}, A^*_{\bar{\iota}_m})$, $\frac{\partial}{\partial s}G(s, A) > 0$ for all $s \in O(s)$.

(b) There exists $\hat{A} > \bar{A}_{\bar{\iota}_m}$ such that for all $A < \hat{A}$, $\frac{\partial}{\partial s}G[\bar{s}(\bar{\iota}_m, A), A] < 0$.

By (a), for $A \in [\hat{\hat{A}}, A^*_{\bar{\iota}_m})$, it is optimal to set s a high as possible within $O(s)$ and hence optimal $s = \bar{s}(\bar{\iota}_m, A)$. This proves Theorem 4.2 (2). By (b), for $A < \hat{A}$, it is optimal to set s below $\bar{s}(\bar{\iota}_m, A)$, and this proves Theorem 4.2 (3). Note also that from (27), for $A \leq \bar{A}_{\bar{\iota}_m}$, $\bar{s}(\bar{\iota}_m, A) = \bar{\iota}_m$.

To prove (a) and (b), we first compute $\frac{\partial}{\partial s}G(s,A)$:

$$\begin{aligned}\frac{\partial}{\partial s}G(s,A) &= \sigma[u'(Q(s))-1]Q'(s)+\sigma_u[v'(Y(\bar{\iota}_m,s))-1]\frac{\partial}{\partial s}Y(\bar{\iota}_m,s)\\ &\quad + n\left\{A-(1+r)\psi'[K(s,A)]\right\}\frac{\partial}{\partial s}K(s,A).\end{aligned} \tag{50}$$

By (F1), $Q'(s) < 0$ and $\frac{\partial}{\partial s}Y(\bar{\iota}_m,s) > 0$, and by (F2), $\frac{\partial}{\partial s}K(s,A) > 0$ and $A-(1+r)\psi'[K(s,A)] < 0$. Thus, the first and the third terms in (50) are non-positive but the second term is non-negative.

Now we prove (a). We use continuity of $\frac{\partial}{\partial s}G(s,A)$ and $\overline{s}(\bar{\iota}_m,A)$. Note that $\overline{s}(\bar{\iota}_m,A)$ converges to zero as A converges to $A^*_{\bar{\iota}_m}$ and is continuous for $A < A^*_{\bar{\iota}_m}$. Since $Q(0)=q^*$ and $K(0,A)=k^*(A)$ but $Y(\bar{\iota}_m,0)<y^*$, $\frac{\partial}{\partial s}G(0,A^*_{\bar{\iota}_m})>0$. Thus, there exists $\tilde{s}>0$ such that $\frac{\partial}{\partial s}G(s,A^*_{\bar{\iota}_m})>0$ for all $s\in[0,\tilde{s}]$. By continuity and compactness, there exists $\tilde{A}_1 < A^*_{\bar{\iota}_m}$ such that $\frac{\partial}{\partial s}G(s,A)>0$ for all $(s,A)\in[0,\tilde{s}]\times[\tilde{A}_1,A^*_{\bar{\iota}_m}]$. Choose $\hat{\hat{A}}\in[\tilde{A}_1,A^*_{\bar{\iota}_m})$ such that $\overline{s}(\bar{\iota}_m,\hat{\hat{A}})\le\tilde{s}$, and this implies that $O(A)\subset[0,\tilde{s}]$ for all $A\in[\hat{\hat{A}},A^*_{\bar{\iota}_m})$.

To prove (b), first note that $\frac{\partial}{\partial s}G[\overline{s}(\bar{\iota}_m,A),A]<0$ for all $A\le\bar{A}_{\bar{\iota}_m}$. This follows immediately from the fact that $\overline{s}(\bar{\iota}_m,A)=\bar{\iota}_m$ for those A's, and that $Q(\bar{\iota}_m)<q^*$ while $Y(\bar{\iota}_m,\bar{\iota}_m)=y^*$. The result then follows from continuity of $\frac{\partial}{\partial s}G(s,A)$ and $\overline{s}(\bar{\iota}_m,A)$.

Part 3 Here we show that optimal $s<r$ for σ_u small. Since $\overline{s}(\bar{\iota}_m,A)$ strictly decreases with A and reaches zero at $A^*_{\bar{\iota}_m}$, there is a unique threshold $\tilde{A}_2$ such that $\overline{s}(\bar{\iota}_m,\tilde{A})=r$. Obviously, for $A>\tilde{A}_2$, optimal $s<r$. Since $\underline{s}(\bar{\iota}_m,A)<r$ for all A, $r\in O(A)$ for all $A\le\tilde{A}_2$. We then show that, for σ_u small, $\frac{\partial}{\partial s}G(s,A)<0$ for all $s\in[r,\bar{\iota}_m]$. Now, we gather the first two terms in (50) at $s=r$, and define

$$\begin{aligned}H(\sigma_u,s) &= \sigma[u'(Q(s))-1]Q'(s)+\sigma_u[v'(Y(\bar{\iota}_m,s))-1]\frac{\partial}{\partial s}Y(\bar{\iota}_m,s)\\ &= \frac{[u'(Q(s))-1]}{u''(Q(s))}-\sigma_u\frac{(1+\bar{\iota}_m)[v'(Y(\bar{\iota}_m,s))-1]}{(1+s)^2v''(Y(\bar{\iota}_m,s))}.\end{aligned} \tag{51}$$

Since both u and v are strictly concave, $H(0,s)<0$ for all $s\in[r,\bar{\iota}_m]$ and hence, by continuity and compactness, there exists $\bar{\sigma}_u$ such that for all $(\sigma_u,s)\in[0,\bar{\sigma}_u]\times[r,\bar{\iota}_m]$, $H(\sigma_u,s)<0$. This implies that, for all $A\le\tilde{A}_2$, $\frac{\partial}{\partial s}G(s,A)<0$ for all $(\sigma_u,s)\in[0,\bar{\sigma}_u]\times[r,\bar{\iota}_m]$. Note that the third term in (50) is negative in the relevant range.

Part 4 Finally, we show that optimal ι increases with A, or, equivalently, optimal s decreases with A. Since $s\ge 0$ always holds, we only need to show this for $A<A^*_{\bar{\iota}_m}$. By Theorem 2.8.3 in Topkis (1998), we only need to show that $O(A)$ decreases with A and $\frac{\partial^2}{\partial A\partial s}G(s,A)<0$. For the former, it follows from that $K(s,A)$ increases with A, and hence both $\underline{s}(\bar{\iota}_m,A)$ and $\overline{s}(\bar{\iota}_m,A)$ decreases with A (see also Figure 3). For the latter, differentiating (50) w.r.t. A, we obtain

$$\frac{\partial^2}{\partial A \partial s} G(s,A) \;=\; n\left\{-\rho A \frac{\partial}{\partial A} K(s,A) - \rho s A \frac{\partial^2}{\partial A \partial s} K(s,A) + \frac{\partial}{\partial s} K(s,A)\right\}. \tag{52}$$

Now, since $\psi(k) = (k+\lambda)^x/(\lambda^{x-1}x) - \lambda/x$, $K(s,A) = \lambda\{[A(1+\rho s)]/(1+r)\}^{\frac{1}{x-1}} - \lambda$. Then,

$$\begin{aligned}
\frac{\partial}{\partial A} K(s,A) &= \frac{\lambda\,[A(1+\rho s)]^{\frac{2-x}{x-1}}}{(x-1)(1+r)^{\frac{1}{x-1}}}(1+\rho s) = \frac{\lambda A^{\frac{2-x}{x-1}}}{(x-1)(1+r)^{\frac{1}{x-1}}}(1+\rho s)^{\frac{1}{x-1}}, \\
\frac{\partial}{\partial s} K(s,A) &= \frac{\lambda\,[A(1+\rho s)]^{\frac{2-x}{x-1}}}{(x-1)(1+r)^{\frac{1}{x-1}}}(A\rho),\;\; \frac{\partial^2}{\partial A \partial s} K(s,A) = \frac{\lambda\rho\,[A(1+\rho s)]^{\frac{2-x}{x-1}}}{(x-1)^2(1+r)^{\frac{1}{x-1}}} > 0.
\end{aligned}$$

Now, for any $s > 0$,

$$-\rho A \frac{\partial}{\partial A} K(s,A) + \frac{\partial}{\partial s} K(s,A) = -\frac{\lambda\,[A(1+\rho s)]^{\frac{2-x}{x-1}}}{(x-1)(1+r)^{\frac{1}{x-1}}}(-\rho^2 A s) < 0.$$

This inequality, together with $\frac{\partial^2}{\partial A \partial s} K(s,A) > 0$ and (52), implies that $\frac{\partial^2}{\partial A \partial s} G(s,A) < 0$. Finally, since the inequality is strict, it also follows that any selection of the optimal s is decreasing (by Theorem 2.8.4 in Topkis (1998)). $\square$

Proof of Lemma 5.1

Proof. **Part 1** Here we prove necessity. Note that, given a portfolio (b,m), in the DM money/deposit market what really matters is the total "wealth" $w = b + \chi m$. If $\chi > (1+\pi)/\varphi$, then, for any targeted wealth w in that market, the problem

$$\min_{(b,m)} \varphi b + (1+\pi)m \text{ s.t. } b + \chi m = w,$$

has a unique solution $m = w/\chi$ and $b = 0$. Similarly, if $\chi < (1+\pi)/\varphi$, the unique solution is $b = w$ and $m = 0$. Thus, to implement both $m > 0$ and $b > 0$ in the CM, it is necessary that $\chi = (1+\pi)/\varphi$. The necessity of (32) follows directly from consumer budget constraint, liquidity constraint, and seller participation constraints.

Following the proposed allocation, the continuation payoff to a consumer in the CM (note that the CM value function is linear in asset holdings and we omit the constant terms) is

$$\begin{aligned}
& -\varphi b^h-(1+\pi)m^h+\beta\left\{\sigma_u[v(y)+m^i-p+b^i]\right\}+\beta\left\{\sigma[u(q)+(b^f-d)]+(1-\sigma-\sigma_u)b^f\right\} \\
= & \ \beta\left\{\sigma[u(q)-d]+\sigma_u[v(y)-p]-\frac{\varphi}{\beta}b^h-\frac{1+\pi}{\beta}m^h+(1-\sigma_u)b^f+\sigma_u(b^i+m^i)\right\} \\
= & \ \beta\left\{\sigma[u(q)-d]+\sigma_u[v(y)-p]-(1+s)b^h-(1+\iota_m)m^h+b^h+m^h\right\} \\
= & \ \beta\left\{-sb^h-\iota_m m^h+\sigma[u(q)-d]+\sigma_u[v(y)-p]\right\}, \qquad (53)
\end{aligned}$$

where the second equality used $s=\varphi/\beta-1$ and $\chi=(1+\pi)/\varphi-1$, and (29). This continuation value has to be non-negative, for otherwise the consumer can deviate to hold nothing and receive no trade. This implies (30).

Now consider (31). First, a consumer can hold just m^i real balances to leave the CM and only participate the DM informal trade, and, in case there is no such match, trade all money into deposits. To deter this deviation, the continuation value from it, which is the left-side of the following equality, must be lower than the one given by (53), reproduced in the right-side:

$$\begin{aligned}
& -(1+\pi)m^i+\beta\{\sigma_u[v(y)+m^i-p]+(1-\sigma_u)\chi m^i\} \\
& =\beta\{-(1+\iota_m)m^i+\sigma_u[v(y)+m^i-p]+(1-\sigma_u)\chi m^i\} \\
& \leq\beta\left\{-sb^h-\iota_m m^h+\sigma[u(q)-d]+\sigma_u[v(y)-p]\right\} \\
\text{iff} \quad & -\iota_m m^i+(1-\sigma_u)(\chi-1)m^i\leq -sb^h-\iota_m m^h+\sigma[u(q)-d] \\
\text{iff} \quad & (\chi-1-\iota_m)m^i-(\chi-1)\sigma_u m^i+sb^h+\iota_m m^h\leq\sigma[u(q)-d] \\
\text{iff} \quad & (\chi-1-\iota_m)m^i-(\chi-1)m^h+\iota_m m^h+sb^h\leq\sigma[u(q)-d] \\
\text{iff} \quad & (\chi-1-\iota_m)(m^i-m^h)+sb^h=s[-\chi(m^i-m^h)+b^h]=sb^i\leq\sigma[u(q)-d],
\end{aligned}$$

where we used $b^h+\chi m^h=b^i+\chi m^i$, $s\chi=1+\iota_m-\chi$, and $m^h=\sigma_u m^i$, and this implies the first condition in (31). For the second condition, consider an informal in the DM money/deposit market. He can deviate to hold deposits only, and, to deter this deviation, we need

$$v(y)+m^i+b^i-p \quad \geq \quad b^h+\chi m^h=b^i+\chi m^i,$$

which is equivalent to the second inequality in (31). Finally, the conditions in (33) are necessary to satisfy the pairwise core requirements.

Part 2 Here we prove sufficiency. Consider the following trading mechanisms. First, $o^d(b+$

m) is the solution to

$$\max_{(q',d'),d'\leq b+m} -q'+d' \text{ s.t. } u(q')-d' \geq \mathbb{I}_{b\geq b^f}[u(q)-d], \tag{54}$$

where $\mathbb{I}_{b\geq b^f}=1$ if $b\geq b^f$ and is zero otherwise. When there are multiple solutions, pick the one with the smallest b'. Similarly, $o^m(m)$ is the solution to

$$\max_{(y',p'),y'\leq m} -y'+p' \text{ s.t. } v(y')-p' \geq \mathbb{I}_{m\geq m^i}[v(y)-p]. \tag{55}$$

Note that, by (33), $o^d(b^f)=(q,d)$ and $o^m(m^i)=(y,p)$. Moreover, given the trading mechanisms, it is straightforward to show that both the consumers and sellers are willing to participate the trades, and, by (33), it is optimal not to renegotiate.

We use $V^f(b,m)$ to denote the continuation value for a formal consumer, and $V^i(b,m)$ for an informal consumer. Now we claim that

$$V^f(b,m) = \begin{cases} \frac{\sigma}{1-\sigma_u}[u(q)-d]+b+\chi m \text{ if } b+\chi m\geq b^f; \\ b+\chi m \text{ otherwise;} \end{cases} \tag{56}$$

$$V^i(b,m) = \begin{cases} [v(y)-p]+b+\chi(m-m^i)+m^i \text{ if } b+\chi m\geq \chi m^i; \\ b+\chi m \text{ otherwise.} \end{cases} \tag{57}$$

First consider (56). Since $\chi\geq 1$, and since (54) implies that the consumer's surplus is constant whenever $b+m\geq b^f$, the consumer is better off holding deposits only. When $w<b^f$, the consumer cannot afford any portfolio that satisfy $b'+m'\geq b^f$, and (54) implies that he will receive zero surplus, it follows that it is optimal to sell all his money holdings for deposits as $\chi\geq 1$. Similar reasoning holds for (57), and we need to use the second inequality in (31) and he only needs χm^i wealth to follow the equilibrium behaviour.

Now we show that (b^h,m^h) maximizes

$$-\varphi b-(1+\pi)m+\beta[\sigma_u V^i(b,m)+(1-\sigma_u)V^f(b,m)]. \tag{58}$$

We show this by considering three thresholds for the portfolio (b,m): (a) (b,m) such that $b+\chi m\geq b^h+\chi m^h$; (b) (b,m) such that $\chi m^i\leq b+\chi m<b^h+\chi m^h$; (c) (b,m) such that $0\leq b+\chi m<\chi m^i$. We claim that for portfolio in range (a), it is optimal to hold $(b,m)=(b^h,m^h)$; for portfolio in range (b), $(b,m)=(0,m^i)$; for (c), $(b,m)=(0,0)$.

For (a), letting $w = b + \chi m$, we have

$$\begin{aligned}
& -\varphi b - (1+\pi)m + \beta[\sigma_u V^i(b,m) + (1-\sigma_u)V^f(b,m)] \\
= \; & -\varphi w + \beta\left\{\sigma_u[v(y) - p + w - \chi m^i] + (1-\sigma_u)\left[\frac{\sigma}{1-\sigma_u}[u(q) - d + w] + \frac{1-\sigma-\sigma_u}{1-\sigma_u}w\right]\right\} \\
= \; & \beta\left\{-sw + \sigma[u(q) - d] + \sigma_u[v(y) - p] - (\chi - 1)m^h\right\},
\end{aligned}$$

which is strictly decreasing in w and hence it is optimal to have $w = b^h + \chi m^h$.

For (b), we have

$$\begin{aligned}
& -\varphi b - (1+\pi)m + \beta[\sigma_u V^i(b,m) + (1-\sigma_u)V^f(b,m)] \\
= \; & -\varphi w + \beta\left\{\sigma_u[v(y) - p + w - \chi m^i] + (1-\sigma_u)w\right\} \\
= \; & \beta\left\{-sw + \sigma_u[v(y) - p - (\chi-1)m^i]\right\},
\end{aligned}$$

which is strictly decreasing in w and hence it is optimal to have $w = \chi m^i$.

Finally, for (c), we have

$$-\varphi b - (1+\pi)m + \beta[\sigma_u V^i(b,m) + (1-\sigma_u)V^f(b,m)] = -\varphi w + \beta w = -\beta s w,$$

which is strictly decreasing in w and hence it is optimal to have $w = 0$.

Thus, we only need to show that

$$\begin{aligned}
& -\varphi b^h - (1+\pi)m^h + \beta[\sigma_u V^i(b^h,m^h) + (1-\sigma_u)V^f(b^h,m^h)] \qquad (59) \\
& \geq -(1+\pi)m^i + \beta[\sigma_u V^i(0,m^i) + (1-\sigma_u)V^f(0,m^i)], \\
& -\varphi b^h - (1+\pi)m^h + \beta[\sigma_u V^i(b^h,m^h) + (1-\sigma_u)V^f(b^h,m^h)] \geq 0. \qquad (60)
\end{aligned}$$

Now, (59) follows from the first condition in (31) and (60) follows from (30). □

Proof of Lemma 5.2

Proof. **<u>Part 1</u>** Here we prove necessity, by considering the deviation to private equity with profit Π_A^*. We need, however, to compute the profit from the proposed allocation. Consider a bank operating in period-t CM. Note that z stands for the amount the government pays to each bank in terms of reserves (which include interest on reserves) in period $t+1$. Regardless of the policy or regulation specified in the proposed allocation, we can summarise the transfer to the bank by L and let $z - L$ be the amount each bank pays to the government in period-t

CM, following the proposed allocation. Thus, the bank's profit can be written as

$$\begin{aligned} & -z-\psi(k)+\varphi b+L+\beta(-b+Ak+z) \\ = \ & \beta\{sb-rz+(1+r)L+Ak-(1+r)\psi(k)\}. \end{aligned} \tag{61}$$

We can give an upper bound on L by the government budget constraint. Since the government does not produce, L has to come from consumer production or to compensate consumers, either through inflation tax ($L > 0$) or to finance deflation ($L < 0$). The total proceeds (or expense) from those each period is given by πm^h, in per consumer terms. Since there are n banks per consumer, we have $nL \leq \pi m^h$. Plugging the maximum L and $b = \rho Ak + z$ into (61) and it has to be no less than Π_A^*, we obtain (35).

<u>Part 2</u> For sufficiency, we first consider the inflation scheme and have the following lemma.

Lemma 6.1. *Let (π, s, m^h) be given with $\pi > 0$. An outcome (k, z, b) that satisfies (a) $b = \rho Ak + z$, (b) the condition (35) at equality, and (c) $k^*(A) \leq k \leq K(s, A)$, can be implemented without explicit taxation.*

Proof. Let $\bar{z}$ denote fixed reserve requirement, η the proportional reserve requirement that is proportional to a bank's capital holding, and ι the interest rate paid on excess reserves, $z - \bar{z} - \eta k$. Hence, a bank profit under portfolio (k', z') with $b' = \rho Ak' + z'$ is given by

$$\Pi(k', z'; \varphi, A) = \beta\left\{(1+s\rho)Ak' - (1+r)\psi(k') - (\iota_m - s)(\bar{z}+\eta k') - \left(\frac{\iota_m - \iota}{1+\iota} - s\right)(z' - \bar{z} - \eta k')\right\}.$$

The FOC to maximize profit is then given by

$$(1+s\rho)A - (\iota_m - s)\eta + \left(\frac{\iota_m - \iota}{1+\iota} - s\right)\eta \ = \ (1+r)\psi'(k'), \ \left(\frac{\iota_m - \iota}{1+\iota} - s\right) = 0. \tag{62}$$

Now, set

$$\eta \ = \ \frac{1}{\iota_m - s}[(1+s\rho)A - (1+r)\psi'(k)], \ \iota = \frac{\iota_m - s}{1+s}. \tag{63}$$

Note that $\eta \geq 0$ since $k \leq K(s, A)$. From (62), this construction ensures that it is optimal for banks for hold $k' = k$, and $z' = z$. Now, budget balance of the central bank requires

$$\frac{\iota}{1+\iota}(z - \bar{z} - \eta k) \leq \pi\left[m^h + (\bar{z}+\eta k) + \frac{z - \bar{z} - \eta k}{1+\iota}\right].$$

taking ι by (63), we have

$$(r-s)z \leq (\iota_m - r)\frac{m^h}{n} + (\iota_m - s)(\bar{z} + \eta k). \tag{64}$$

We choose

$$\bar{z} = \frac{(r-s)}{(\iota_m - s)}z - \frac{(\iota_m - r)}{(\iota_m - s)}\frac{m^h}{n} - \eta k. \tag{65}$$

which ensures (64). Now we show that $\bar{z} \geq 0$:

$$\begin{aligned}
\bar{z} &= \frac{(r-s)}{(\iota_m - s)}z - \frac{(\iota_m - r)}{(\iota_m - s)}\frac{m^h}{n} - \eta k \\
&= \frac{(r-s)}{(\iota_m - s)}(b - \rho A k) - \frac{(\iota_m - r)}{(\iota_m - s)}\frac{m^h}{n} - \frac{1}{\iota_m - s}[(1+s\rho)A - (1+r)\psi'(k)]k \\
&= \frac{1}{(\iota_m - s)}\left\{(r-s)b - (r-s)\rho A k - (1+s\rho)Ak + (1+r)\psi'(k)k - \frac{(1+r)\pi m^h}{n}\right\} \\
&= \frac{1}{(\iota_m - s)}\left\{rz - sb - Ak + (1+r)\psi'(k)k - \frac{(1+r)\pi m^h}{n}\right\} \\
&= \frac{1}{(\iota_m - s)}\left\{-(1+r)\psi(k) + (1+r)\psi'(k)k - (1+r)\Pi_A^*\right\} \geq 0,
\end{aligned} \tag{66}$$

where the second equality uses $z = b - \rho A k$ and the construction of η, the third uses $\iota_m - r = (1+r)\pi$, the fourth uses $z = b - \rho A k$, the fifth uses (35) at equality, and the last inequality follows from the convexity of ψ and hence $\psi'(k')k' - \psi(k')$ is strictly increasing and $k \geq k^*(A)$ (recall (34)).

Finally, we only need to make sure that each bank's profit following the proposed allocation is no less than Π_A^*. The profit is given by

$$\begin{aligned}
&\beta\left\{(1+s\rho)Ak - (1+r)\psi(k) - (\iota_m - s)(\bar{z} + \eta k)\right\} \\
\geq\ &\beta\left\{(1+s\rho)Ak - (1+r)\psi(k) - (\iota_m - s)\eta k - (r-s)z + (\iota_m - r)\frac{m^h}{n} + (\iota_m - s)\eta k\right\} \\
=\ &\beta\left\{(1+s\rho)Ak - (1+r)\psi(k) - (r-s)z + (\pi(1+r))\frac{m^h}{n}\right\} = \Pi_A^*.
\end{aligned}$$

where the last inequality follows from (35) at equality. □

Next, we show that any outcome (k, z, b) with $b = \rho A k + z$ that satisfies (35) can be implemented with taxation, no matter what π is. Budget constraint requires

$$\frac{\iota}{1+\iota}z \leq \pi\left(\frac{m^h}{n} + \frac{z}{1+\iota}\right) + \tau_0 + \tau_1 k. \tag{67}$$

The bank profit with portfolio (k', z') and $b' = \rho Ak' + z'$ is given by

$$\Pi(k, z; \varphi, A) = \beta \left\{ (1 + s\rho)Ak - (1 + r)\psi(k) - (1 + r)\tau_0 - (1 + r)\tau_1 k + \left(\frac{\iota_m - \iota}{1 + \iota} - s \right) z \right\}.$$

Now, set ι as in (63), and set

$$\tau_1 = \frac{1}{1+r}[(1 + \rho s)A - (1 + r)\psi'(k)], \tag{68}$$

$$\tau_0 = \max\left\{0, \frac{\iota - \pi}{1 + \iota} z - \pi \frac{m^h}{n} - \tau_1 k \right\} = \max\left\{0, \frac{r - s}{1 + r} z - \pi \frac{m^h}{n} - \tau_1 k \right\}, \tag{69}$$

where the second equality in (69) follows from (63). (68) ensure that the banks hold $k' = k$ and $z' = z$ according to the FOC's, and (69) ensures (67). Note that τ_1 will be negative if $k > K(s, A)$. We also need to ensure that banks make profits no less than Π^*_A:

$$\begin{aligned} \Pi(k, z; \varphi, A) &= \beta \{(1 + s\rho)Ak - (1 + r)\psi(k) - (1 + r)\tau_0 - (1 + r)\tau_1 k\} \\ &\geq \beta \left\{ (1 + s\rho)Ak - (1 + r)\psi(k) - (r - s)z + (1 + r)\pi \frac{m^h}{n} \right\} \geq \Pi^*_A, \end{aligned}$$

where the last inequality follows from (35) and $b = \rho Ak + z$. □

Proof of Theorem 5.1

Proof. **Part 1** Here we show necessity. Note that from the proof of Lemma 5.2, (35) is necessary even if $b < n\rho Ak + nz$. From (30), (35), and market clearing, we have

$$\begin{aligned} -sb^h - \iota_m m^h + \sigma_u[v(y) - p] + \sigma[u(q) - d] &\geq 0, \\ n[-(r - s)z + (1 + \rho s)Ak - (1 + r)\psi(k)] + (\iota_m - r)m^h &\geq n(1 + r)\Pi^*_A, \\ sb^h = sn(\rho Ak + z). \end{aligned}$$

Now, adding the three together, we obtain (note that $\chi = (1 + \iota_m)/(1 + s)$)

$$n[-rz + Ak - (1 + r)\psi(k)] + \sigma_u[v(y) - p] + \sigma[u(q) - d] - rm^h \geq n(1 + r)\Pi^*_A,$$

which implies (36) as $p \geq y$ and $d \geq q$.

Now, for (37), from (31)-(32), we have (note, again, that $\chi = (1 + \iota_m)/(1 + s)$)

$$q \leq b^f = b^h + \chi m^h = n(\rho Ak + z) + \chi m^h, \tag{70}$$

$$v(y) - p - (\chi - 1)m^i \geq 0. \tag{71}$$

Now, adding (70) with (71) multiplied by σ_u, we have

$$\sigma_u[v(y) - p] + n(\rho Ak + z) + \chi m^h \geq q + (\chi - 1)(\sigma_u m^i) = q + (\chi - 1)m^h,$$

where the last inequality follows from $m^h = \sigma_u m^i$, and this implies (37) since $y \leq p$.
Part 2 Here we prove sufficiency. We need to reconstruct the policy variables. Set

$$\begin{aligned} b^h &= n(\rho Ak + z),\ m^i = y = m^h/\sigma_u,\ d = q,\ p = y. && (72) \\ \chi &= \frac{\sigma_u[v(y) - y]}{m^h} + 1 > 1,\ s = \frac{\sigma[u(q) - q]}{b^h + \chi m^h},\ \iota_m = \chi(1 + s) - 1 > s. && (73) \end{aligned}$$

Note that these constructions imply that

$$\begin{aligned} -s(b^h + \chi m^h) + \sigma[u(q) - q] &= 0, && (74) \\ v(y) - y - (\chi - 1)m^i &= 0, && (75) \end{aligned}$$

which imply (30); note that $d = q$ and $p = y$. Moreover, since $b^i < b^h + \chi m^h$, this implies that both conditions in (31) are met. We also need to show that $b^i \geq 0$. Since

$$\begin{aligned} b^i &= b^h + \chi m^h - \chi m^i = n(\rho Ak + z) + \left[\frac{\sigma_u[v(y) - y]}{m^h} + 1\right]\left(m^h - \frac{m^h}{\sigma_u}\right) \\ &= n(\rho Ak + z) - (1 - \sigma_u)[v(y) - y] - \frac{m^h}{\sigma_u} + m^h. \end{aligned}$$

Thus, $b^i \geq 0$ if and only if

$$n(\rho Ak + z) + \sigma_u[v(y) - y] + m^h \geq [v(y) - y] + \frac{m^h}{\sigma_u}. \tag{76}$$

Now, by (37), (76) holds if $q \geq v(y)$ and if $y = m^h/\sigma_u$, which hold by conditions (a) and (b).

Since $y = p = m^i$, (33) is satisfied if $q = q^*$. Otherwise, (37) holds with equality by condition (c) and

$$b^h + \chi m^h = n(\rho Ak + z) + \left(\frac{\sigma_u[v(y) - y]}{m^h} + 1\right)m^h = n(\rho Ak + z) + \sigma_u[v(y) - y] + m^h = q = d.$$

Hence, (33) is satisfied.

Now we move to the banks. Since (72) implies that $b = b^h/n = \rho Ak + z$, we only need to verify (35):

$$
\begin{aligned}
& sb - rz + Ak - (1+r)\psi(k) + \frac{(1+r)\pi m^h}{n} \\
= \; & \frac{\sigma[u(q)-q] - s\chi\sigma_u y}{n} - rz + Ak - (1+r)\psi(k) + \frac{(\iota_m - r)\sigma_u y}{n} \\
= \; & \frac{1}{n}\left\{\sigma[u(q)-q] - s\frac{v(y)}{y}\sigma_u y - rz + Ak - (1+r)\psi(k) + (\iota_m - r)\sigma_u y\right\} \\
= \; & \frac{1}{n}\left\{\sigma[u(q)-q] - rz + Ak - (1+r)\psi(k) + \sigma_u[v(y) - (1+r)y]\right\} \geq (1+r)\Pi_A^*.
\end{aligned}
$$

where the first equality follows from $nb = b^h$ and (74) and $m^h = \sigma_u y$ and $(1+r)\pi = \iota_m - r$, the second from $\chi = v(y)/y$, the third from $\iota_m y - sv(y) = v(y) - y$, and the last inequality follows from (36) and $m^h = \sigma_u y$. □

Proof of Theorem 5.2

Proof. **Part 1** Here we prove (1) by constructing A^{**}, and characterize the constrained efficient allocation by FOC's for $A < A^{**}$.

Part 1 (i) (Construction of A^{**}) We consider the following condition

$$n[(1+r\rho)Ak - (1+r)\psi(k)] - rq + \sigma[u(q)-q] + (1+r)\sigma_u[v(y)-y] \geq n(1+r)\Pi_A^*, \quad (77)$$

which is obtained from summing up (36) and (37) multiplied by r, and hence this condition is necessary for implementation by Theorem 5.1. Let A^{**} be the smallest $A \geq 0$ such that

$$r[n\rho A k^*(A) - q^* - \sigma_u v(y^*)] + \sigma_u[v(y^*) - (1+r)y^*] + \sigma[u(q^*) - q^*] \geq 0. \quad (78)$$

Clearly $A^{**} < A^*$. Note that for all $A \geq A^{**}$, the allocation $(q^e, k^e, y^e) = (q^*, k^*(A), y^*)$ satisfies (77).

Now we show that for such A's, $(q^*, k^*(A), y^*)$ is implementable using Theorem 5.1. Let

$$z = \max\left\{0, \frac{q^*}{n} - \rho A k^*(A) - \frac{1}{n}\sigma_u v(y^*)\right\},\; m^h = \sigma_u y^*. \quad (79)$$

By construction (37) and (38) are satisfied. Thus, by Theorem 5.1, we only need to show (36), and the additional conditions, among those we need only to verify $q^* \geq v(y^*)$.

To verify $q^* \geq v(y^*)$, first note that by (2), $v'(y^*) = \theta u'(y^*) = 1$. Now, let q^r be the unique $q > 0$ that solves

$$u(q) = \frac{\sigma + r}{\sigma} q. \quad (80)$$

By (39), $q^* > q^r$. Moreover, $q^r > Q(r)$, and, by (43),

$$q^* > q^r = \frac{\sigma}{\sigma + r} u(q^r) > \frac{\sigma}{\sigma + r} u[Q(r)] > \theta u(y^*) = v(y^*). \tag{81}$$

Finally, we prove (36) for $\langle q^*, k^*(A), y^*, z, m^h \rangle$, which now reads

$$\sigma[u(q^*) - q^*] + \sigma_u[v(y^*) - (1 + r)y^*] \geq 0, \tag{82}$$

since $(1 + r)\Pi_A^* = Ak^*(A) - (1 + r)\psi[k^*(A)]$ and $m^h = \sigma_u y^*$. But, by (81),

$$\sigma[u(q^*) - q^*] > \sigma[u(q^r) - q^r] = rq^r > rv(y^*) > \sigma_u r y^*.$$

Part 1 (ii) (Characterization of optimal allocation for $A < A^{**}$)

Now suppose that $A < A^{**}$. Consider the following maximization problem:

$$\max_{(q,k,y,z)} \sigma[u(q) - q] + n[Ak - (1 + r)\psi(k)] + \sigma_u[v(y) - y] \tag{83}$$

subject to

$$\begin{aligned} -rnz + n[Ak - (1 + r)\psi(k)] + \sigma[u(q) - q] + \sigma_u[v(y) - (1 + r)y] \ &\geq \ n(1 + r)\Pi_A^*, && (84) \\ n\rho Ak + nz + \sigma_u v(y) \ &\geq \ q, && (85) \\ z \ &\geq \ 0. && (86) \end{aligned}$$

The objective function is the same as (3). The constraint (84) is derived from (36) with $m^h = \sigma_u y$, and (85) is derived from (37) with $m^h = \sigma_u y$. Since the objective function is strictly concave and the constraints convex, there is a unique solution, denoted by (q^e, k^e, y^e); the solution z^e may not be unique in case the constraints are relaxed, but, as we will see below, they will not be.

The solutions can then be characterized by FOC's. Let λ_1 be the lagrangian multiplier for (84), λ_2 for (85), and λ_3 for (86). Then, the FOCs are given by

$$\begin{aligned} \sigma[u'(q) - 1] + \lambda_1 \sigma[u'(q) - 1] - \lambda_2 \ &= \ 0, && (87) \\ A - (1 + r)\psi'(k) + \lambda_1[A - (1 + r)\psi'(k)] + \lambda_2 \rho A \ &= \ 0, && (88) \\ \sigma_u[v'(y) - 1] + \lambda_1 \sigma_u[v'(y) - (1 + r)] + \lambda_2 \sigma_u v'(y) \ &= \ 0, && (89) \\ -rn\lambda_1 + n\lambda_2 + \lambda_3 \ &= \ 0. && (90) \end{aligned}$$

The FOC's then imply

$$\sigma[u'(q^e) - 1] = \frac{\lambda_2}{1+\lambda_1} = \frac{(1+r)\psi'(k^e) - A}{\rho A}, \tag{91}$$

$$v'(y^e) = \frac{1+\lambda_1(1+r)}{1+\lambda_1+\lambda_2}. \tag{92}$$

We first show that if $\lambda_2 = 0$, then $(q^e, k^e, y^e) = (q^*, k^*, y^*)$. Clearly, (91) implies that $q^e = q^*$ and $k^e = k^*$. If $\lambda_1 = 0$, then (92) implies $y^e = y^*$. Otherwise, (92) implies that $y^e < y^*$ and (90) implies that $\lambda_3 > 0$. This then implies that both (86) is binding and hence $q^* \le \rho A k^* + \sigma_u v(y^e)$. However, since $A < A^{**}$, by (82), $q^* > \rho A k^* + \sigma_u v(y^*)$, this leads to a contradiction. Finally, since (q^*, k^*, y^*) satisfies (84) and (85) for some z only if (77) holds, it follows that (q^*, k^*, y^*) cannot be a solution for $A < A^{**}$ and hence $\lambda_2 > 0$. Moreover, (90) implies that $\lambda_1 > 0$ as well.

Thus, the solution satisfies $\lambda_1 > 0$ $\lambda_2 > 0$, $q^e < q^*$ and $k^e > k^*(A)$ and that it satisfies (84) and (85) at equality. Finally, by (90), $\lambda_2 \le r\lambda_1$ and hence (92) implies that $y^e < y^*$.

Now, we claim that

$$-rq^e + \sigma[u(q^e) - q^e] < 0, \tag{93}$$

and hence $q^r < q^e < q^*$. Indeed, since the constraints are binding, (77) is satisfied by (q^e, k^e, y^e) at equality. If (93) does not hold, then (77) (q^e, k^e, y^e) at equality implies that $k^e < k^*$, a contradiction.

Now we show that (q^e, k^e, y^e, z^e) is implementable. Since (85) is binding, (37) holds with equality with $m^h = \sigma_u y^e$. Moreover, $q^e > q^r > v(y^e) = v(y^*)$ by (81). These ensure conditions (a)-(c) in Theorem 5.1, and we only need to show (36). For (36),

$$\begin{aligned}
& -r(nz + m^h) + n[Ak^e - (1+r)\psi(k^e)] + \sigma_u[v(y^e) - y^e] + \sigma[u(q^e) - q^e] \\
= \; & -rq^e + n\rho A k^e + r\sigma_u[v(y^e) - y^e] + n[Ak^e - (1+r)\psi(k^e)] + \sigma_u[v(y^e) - y^e] + \sigma[u(q^e) - q^e] \\
= \; & n[(1+r\rho)Ak^e - (1+r)\psi(k^e)] + (1+r)\sigma_u[v(y^e) - y^e] - rq^e + \sigma[u(q^e) - q^e] = n(1+r)\Pi_A^*,
\end{aligned}$$

where the last equality follows from (77) at equality.

Part 2 Here we show (2) and (3), taking the constrained efficient allocation (q^e, k^e, y^e, z^e) from Part I as given. We construct $\hat{A}$ according to whether (40) holds or not. We first show optimal $\pi > 0$ for $A < \hat{A}$ and finally show optimal $\pi < 0$ for $A > \hat{A}$.

Part 2 (i) Here we construct $\hat{A}$, show that optimal $\pi > 0$ for all $A < \hat{A}$, assuming that (40) holds.

We will construct π and equilibrium $(d, p, b^h, m^h, \chi, s)$ according to (72) and (73), unless otherwise specified. For $A > A^{**}$, we construction z according to (79), and denote it by

$z^e(A)$. Since at $A = A^{**}$, (78) holds with equality and since (82) holds with strict inequality, at $A = A^{**}$, $n\rho Ak^*(A) + \sigma_u v(y^*) < q^*$. Thus, $z^e(A) > 0$ for a range of $A > A^{**}$ and is strictly decreasing in A. Moreover, at $A = A^{**}$,

$$rz^e(A) = \frac{rq^*}{n} - r\rho Ak^* - r\sigma_u \frac{v(y^*)}{n} = \frac{1}{n}\left\{\sigma[u(q^*) - q^*] + \sigma_u[v(y^*) - (1+r)y^*]\right\},$$

since (78) holds with equality. Thus, at $A = A^{**}$, by (40),

$$\frac{v(y^*)}{y^*}\left(\frac{nrz^e(A) - \sigma_u[v(y^*) - (1+r)y^*]}{q^*} + 1\right) = \frac{v(y^*)}{y^*}\left(\frac{\sigma[u(q^*) - q^*]}{q^*} + 1\right) > 1 + r.$$

Now, let $\hat{A}$ be the largest A such that

$$\frac{v(y^*)}{y^*}\left(\frac{rz^e(A) - \frac{\sigma_u}{n}[v(y^*) - (1+r)y^*]}{q^*} + 1\right) \geq 1 + r,$$

and hence $\hat{A} > A^{**}$ as $z^e(A)$ strictly decreases with A.

Now we show that for all $A < \hat{A}$, optimal $\pi > 0$ and optimal implementation is without explicit taxation. For each $A \in [A^{**}, \hat{A})$, we set parameters according to (72) and (73), except for s, which is given by

$$s(A) = \frac{nrz^e(A) - \sigma_u[v(y^*) - (1+r)y^*]}{q^*}. \tag{94}$$

Then, $\pi > 0$ since $A < \hat{A}$. Finally, by construction, it is straightforward to verify that, for each $A \in [A^{**}, \hat{A})$ and $b = \rho Ak^* + z^e(A)$,

$$\begin{aligned} & s(A)b - rz^e(A) + \frac{(\iota_m - r)\sigma_u y^*}{n} \\ = \; & s(A)\frac{q^*}{n} - rz^e(A) + \frac{\sigma_u[v(y^*) - (1+r)y^*]}{n} = 0. \end{aligned}$$

That is, (35) is satisfied with equality. Now, by Lemma 6.1, we can implement this allocation without explicit taxes.

Here we turn to the case $A < A^{**}$. By Part I, $q^e < q^*$ and $y^e \leq y^*$, (40) implies that

$$\frac{v(y^e)}{y^e}\frac{\sigma[u(q^e) - q^e] + q^e}{q^e} > 1 + r, \tag{95}$$

for all $A < A^{**}$ by the concavity of u and v. We use Lemma 6.1 to show that (q^e, k^e, y^e) can be implemented without explicit taxation. First we claim

$$z^e = \frac{q^e}{n} - \rho A k^e - \frac{\sigma_u v(y^e)}{n} > 0 \text{ and } m^h = \sigma_u y^e. \tag{96}$$

To do so, we set parameters according to (72) and (73), and η and $\bar{z}$ given by (63) and (65). First note that, by (93), $s \in (0, r)$. Now, we show that $\eta > 0$. By (73), $\iota_m > s$. By (91) and concavity of u,

$$s = \frac{\sigma[u(q^e) - q^e]}{q^e} > \sigma[u'(q^e) - 1] = \frac{(1+r)\psi'(k^e) - A}{\rho A},$$

which implies that, by (63), $\eta > 0$. Moreover, this implies that $k^e \in (k^*, K(s, A))$, which, by (66), implies also that $\bar{z} > 0$.

Now, by Lemma 6.1, we only need to show (35) holds with equality and $\iota_m > r$. Once these conditions are met, the proof of Lemma 6.1 also implies that $z^e > 0$. First, by construction,

$$1 + \iota_m = \chi(1+s) = \frac{v(y^e)}{y^e} \frac{\sigma[u(q^e) - q^e] + q^e}{q^e} > 1 + r$$

by (95), and hence $\iota_m > r$. Second, for (35),

$$\begin{aligned}
& sb + [Ak^e - (1+r)\psi(k^e)] - rz^e + (1+r)\pi \frac{\sigma_u y^e}{n} \\
= \; & s\left[\frac{q^e}{n} - \frac{\sigma_u v(y^e)}{n}\right] + [Ak^e - (1+r)\psi(k^e)] - r\left[\frac{q^e}{n} - \rho A k^e - \frac{\sigma_u v(y^e)}{n}\right] + (\iota_m - r)\frac{\sigma_u y^e}{n} \\
= \; & [A(1+\rho r)k^e - (1+r)\psi(k^e)] - (r-s)\frac{q^e}{n} + r\frac{\sigma_u[v(y^e) - y^e]}{n} + \frac{\sigma_u[\iota_m y^e - sv(y_e)]}{n} \\
= \; & [A(1+\rho r)k^e - (1+r)\psi(k^e)] - (r-s)\frac{q^e}{n} + \frac{(1+r)\sigma_u[v(y^e) - y^e]}{n} \\
= \; & [A(1+\rho r)k^e - (1+r)\psi(k^e)] + \frac{-rq^e + \sigma[u(q^e) - q^e]}{n} + \frac{\sigma_u(1+r)[v(y^e) - y^e]}{n} \\
= \; & (1+r)\Pi_A^*,
\end{aligned}$$

where the first equality uses $b = \rho A k^e + z^e$ and (96), the third uses $\iota_m y^e - sv(y^e) = v(y^e) - y^e$, the fourth uses (73) to substitute s, and the last uses the fact that (q^e, k^e, y^e) satisfies (77) at equality.

Part 2 (ii) Here we show that optimal s and π decrease with A for $A < A^{**}$, assuming that (40) holds.

Since optimal $\pi > 0$ and $z^e > 0$, we have $\lambda_3 = 0$ in (87)-(90), and hence the allocation (q^e, k^e, y^e) is characterized by $y^e = y^*$, (91), and (77) at equality. Thus, (q^e, k^e), as a function of A, is determined by the implicit functions, $F_1(q,k,A)=0$ and $F_2(q,k,A)=0$, with

$$\begin{aligned}
F_1(q,k,A) &= \rho A\sigma[u'(q)-1]-[(1+r)\psi'(k)-A],\\
F_2(q,k,A) &= -rq+\sigma[u(q)-q]+[(1+\rho r)A-(1+r)\psi'(k)]+(1+r)\sigma_u[v(y^*)-y^*]-(1+r)\Pi_A^*.
\end{aligned}$$

Since $\lambda_3=0$, by (90), $\lambda_2=r\lambda_1$, one can verify using (91) that

$$\begin{aligned}
\frac{\partial}{\partial q}F_1(q^e,k^e,A) &< 0,\ \frac{\partial}{\partial k}F_1(q^e,k^e,A)<0,\\
\frac{\partial}{\partial q}F_2(q^e,k^e,A) &< 0,\ \frac{\partial}{\partial k}F_2(q^e,k^e,A)>0,\\
\frac{\partial}{\partial A}F_1(q^e,k^e,A) &> 0,\ \frac{\partial}{\partial A}F_2(q^e,k^e,A)>0.
\end{aligned}$$

Thus, by the Implicit Function Theorem,

$$\frac{\mathrm{d}}{\mathrm{d}A}q^e = \frac{-\frac{\partial}{\partial A}F_1(q^e,k^e,A)\frac{\partial}{\partial k}F_2(q^e,k^e,A)+\frac{\partial}{\partial A}F_2(q^e,k^e,A)\frac{\partial}{\partial k}F_1(q^e,k^e,A)}{\frac{\partial}{\partial q}F_1(q^e,k^e,A)\frac{\partial}{\partial k}F_2(q^e,k^e,A)-\frac{\partial}{\partial k}F_1(q^e,k^e,A)\frac{\partial}{\partial q}F_2(q^e,k^e,A)}>0. \quad (97)$$

Now, since

$$s=\frac{\sigma[u(q^e)-q^e]}{q^e}$$

and since u is strictly concave, s is strictly decreasing in A. Since $y^e=y^*$, this also implies π decreases with A.

Part 2 (iii) Here we construct $\hat{A}$, assuming that (40) does not hold. The argument for optimal $\pi>0$ and s strictly decreasing in A is similar to the ones before and is omitted.

Since (40) does not hold,

$$\begin{aligned}
& -rq^*+\sigma[u(q^*)-q^*]+\sigma_u[v(y^*)-(1+r)y^*]\\
\leq\ & -\frac{v(y^*)\{\sigma[u(q^*)-q^*]+q^*\}}{y^*}+q^*+\sigma[u(q^*)-q^*]+\sigma_u\left[v(y^*)-\frac{v(y^*)\{\sigma[u(q^*)-q^*]+q^*\}}{q^*}\right]\\
=\ & -\frac{[v(y^*)-y^*]\{\sigma[u(q^*)-q^*]+q^*\}}{y^*}+\sigma_u\left[-\frac{v(y^*)\sigma[u(q^*)-q^*]}{q^*}\right]<0. \qquad (98)
\end{aligned}$$

Since the problem (83)-(86) is continuous in A, its solution is continuous in A as well because of uniqueness by the Theorem of Maximum. Now, when $A=0$, the solution is the unique q that satisfies

$$-rq+\sigma[u(q)-q]+(1+r)\sigma_u[v(y^*)-y^*]=0,$$

a solution denoted by q_0^e. By (98), $q_0^e < q^*$, and $q_0^e > q^r$ defined by (80). Moreover,

$$\begin{aligned} & \frac{v(y^*)}{y^*}\left(\frac{\sigma[u(q_0^e) - q_0^e]}{q_0^e} + 1\right) = \frac{v(y^*)}{y^*}(1+r)\left\{1 - \frac{\sigma_u[v(y^*) - y^*]}{q_0^e}\right\} \\ > \quad & \frac{v(y^*)}{y^*}(1+r)\left\{1 - \frac{\sigma_u[v(y^*) - y^*]}{\sigma_u v(y^*)}\right\} = 1 + r, \end{aligned}$$

where the last inequality uses the fact that $q_0^e > v(y^*)$.

Now, consider the more relaxed problem to maximize (83) but only subject to (77). Denote the optimal solution by $(\tilde{q}, \tilde{k}, \tilde{y})$. Clearly $\tilde{y} = y^*$. Also, when $A = 0$, the solution is $k = 0$ and $q = q_0^e$ and (95) holds. By continuity, there exists $A_0 > 0$ such that for all $A < A_0$,

$$\frac{v(y^*)}{y^*}\left(\frac{\sigma[u(q^e) - q^e]}{q^e} + 1\right) > 1 + r.$$

Now, let $\hat{A}$ be the largest A such that (95) holds with weak inequality for $(\tilde{q}, \tilde{k}, \tilde{y})$. By the same arguments as in (a) above, $\tilde{q}$ is strictly increasing in A. Hence, for all $A < \hat{A}$, (40) holds with strict inequality for $(\tilde{q}, \tilde{k}, \tilde{y})$. By the same arguments as before, $(\tilde{q}, \tilde{k}, \tilde{y}) = (q^e, k^e, y^e)$ and it can be implemented without explicit taxes and $\pi > 0$.

<u>Part 3</u> Here we show that if $A > \hat{A}$, then optimal $\pi < 0$, and if $A < A^{**}$, the optimal $\tau_0 > 0$ and optimal $\tau_1 > 0$. First we show that if $A \in (\hat{A}, A^{**})$, then optimal $\tau_0 > 0$ and optimal $\tau_1 > 0$. We can use exactly the same argument to show that $k^e \in (k^*(A), K(s, A))$, with $s = \sigma[u(q) - q]/q$. Thus, by (68), $\tau_1 > 0$. Similarly, the fact that $\tau_0 > 0$ follows exactly the same argument as in (66).

Finally, we show that optimal $\pi \geq 0$ if $A > \hat{A}$. When $\hat{A} < A^{**}$, we show that if (q^e, k^e, y^e) can be implemented with $\pi > 0$, then (40) holds. First note that since both constraints (84) and (85) are binding, it follows that for any intervention, (30)-(32) must be binding as well, except for the first condition in (31). Thus, the constructions in (72) and (73) are unique. This implies that $\pi > 0$, or, equivalently, $\iota_m > r$, if and only if (40) holds.

Now suppose that $\hat{A} \geq A^{**}$. When looking at (66), the first equality gives an upper bound on $\bar{z}$. When the $k = k^*$, the very last inequality holds with equality. However, when (35) is slack, those computations show that we have $\bar{z} < 0$. Thus, the condition (35) is tight when implementing with inflation for $k = k^*$. As a result, the $s(A)$ constructed in (94), which is designed to keep (35) at equality, is also tight. This implies that, if $A > \hat{A}$, then implementing with $\pi > 0$ and without explicit taxation is impossible. $\square$

Appendix B: Extensions and robustness

B1. Implementable allocations under general instruments

Here we show that by discriminating money and deposits in our trading mechanism with formal sellers does not alter the constrained-efficient allocations, nor does banking fees on consumers. We do so by demonstrating that these additional instruments do not alter the necessary conditions that we identify in Theorem 5.1 that determines the constrained efficient allocations.

Following the literature and as in the main text, we consider a simple game to implement trades in the DM. An outcome for the DM consists of the portfolio of each consumer not in a underground meeting, (b^d, m^d), the portfolio of each in a underground meeting, (b^u, m^u), and the respective trades, (q, d, p_1) and (y, p_2). Note that in the main text we impose $m^d = 0$. Note also that these also determine the portfolio leaving the CM, as that would be symmetric among all consumers.

In pairwise meetings where deposits are transferable, a mechanism is a mapping $o_d(b, m) = (q, d, p_1)$ that maps the consumer's portfolio to a trade (q, d, p_1). The rule of the game is as follows. First, both the consumer and the seller choose to accept the proposed trade according to the mechanism o_d. If both accept it, then the trade is executed; otherwise, the consumer makes a take-it-leave-it offer to the producer. Similarly, in underground meetings, a mechanism is a mapping $o_m(m) = (y, p)$ that maps the consumer's money holding m to a trade (y, p). The rule of the game is as follows. First, both the consumer and the informal seller choose to accept the proposed trade according to the mechanism o_m. If both accept it, then the trade is executed; otherwise, the consumer makes a take-it-leave-it offer to the seller.

In terms of other parameters exogenous to the DM trades, they include banking fees τ, the spread s, and the inflation rate π. We assume that consumers are liable to pay the fee τ when they turn up to purchase deposits or to renew their current holdings. We also assume a strong form of punishment–if the consumer fails to pay the banking fee, he will be excluded from banking services in the future. Of course, a weaker punishment can only strengthen our results.

Lemma 6.2. *For a given (τ, s, π), the outcome $[\chi, (b^h, m^h), (b^d, m^d), (b^u, m^u), y, q]$ is imple-*

mentable for households only if $\chi = (1+\pi)/\varphi$ and the following conditions hold:

$$
\begin{aligned}
-s(b^h + \chi m^h) - (1+r)\tau + \sigma_u[v(y) - p_2] + \sigma[u(q) - d - p_1] - (\chi - 1)m^h &\geq 0, && (99)\\
v(y) - p_2 - (\chi - 1)m^u &\geq 0, && (100)\\
m^d \geq p_1, b^d \geq d,\ d + p_1 \geq q,\ m^u \geq p &\geq y, && (101)\\
(1-\sigma_u)m^d + \sigma^u m^u = m^h,\ b^d + \chi m^d = b^h + \chi m^h = b^u + \chi m^u. & && (102)
\end{aligned}
$$

The proof follows the same outline as that in the main text. The proof for $\chi = (1+\pi)/\varphi$ is exactly the same and is omitted. Condition (99) is necessary because the consumer can simply decide not to purchase any new deposits, to renew any existing one, nor to purchase any money, and leave the CM without any assets and avoid the banking fee. Condition (100) follows exactly the same reasoning as in the main text. Condition (101) simply lays out the liquidity constraints, and (102) consists of market clearing condition for DM market for money and deposits.

Proof. Following the proposed allocation, the continuation payoff to a household in the CM (note that the CM value function is linear in asset holdings and we omit the constant terms) is given by

$$
\begin{aligned}
& -\varphi b^h - (1+\pi)m^h - \tau + \frac{\beta}{1-\beta}\left\{\sigma_u[v(y) + m^u - p_2 + b^u]\right\}\\
+ \quad & \frac{\beta}{1-\beta}\left\{\sigma[u(q) + (b^d + m^d - d - p_1)] + (1-\sigma_u)(b^d + m^d) - \varphi b^h - (1+\pi)m^h - \tau\right\}\\
= \quad & -\varphi(b^h + \chi m^h) - \tau + \frac{1}{r}\left\{\sigma[u(q) - d - p_1] + \sigma_u[v(y) - p_2]\right\}\\
+ \quad & \frac{1}{r}\left\{(1-\sigma_u)(b^d + m^d) + \sigma_u(b^u + m^u) - \varphi(b^h + \chi m^h) - \tau\right\}\\
= \quad & -\varphi(b^h + \chi m^h) - \tau + \frac{1}{r}\left\{\sigma[u(q) - d - p_1] + \sigma_u[v(y) - p_2]\right\}\\
+ \quad & \frac{1}{r}\left\{(1-\varphi)(b^h + \chi m^h) - (\chi - 1)m^h - \tau\right\}\\
= \quad & \frac{1}{r}\left\{-s(b^h + \chi m^h) - (1+r)\tau + \sigma[u(q) - d - p_1] + \sigma_u[v(y) - p_2] - (\chi - 1)m^h\right\},
\end{aligned}
$$

and this has to be non-negative, for otherwise the household can deviate to hold nothing and avoid the fee and receive no trade hereafter. This implies (99). □

Now we turn to the firms, and we give necessary conditions for implementing a proposed outcome, (k, z, b). Here we allow for dynamic incentives as well. Specifically, we allow banks to issue κ unsecured deposits, and the pledgeability constraint would read

$$b \leq \rho Ak + z + \kappa. \tag{103}$$

To ensure banks pay back the unsecured component, κ, we assume that a defaulting bank is excluded from deposit issuance in all future periods (but can still conduct its business with private equity).

Our proof is based on participation, meaning that the deviation we consider is for the bank to opt to private equity and make profit Π_A^*. Therefore, we do not need to specify the interventions. Instead, we only need to compute the maximum profit a bank can make by following the proposed allocation, taken as given parameters that will be jointly determined with consumer allocations, (τ, s, π) and m^h.

Lemma 6.3. *For a given (τ, s, π) and m^h, the outcome (k, z, κ, b) with $b = \rho Ak + z + \kappa$ is implementable for banks only if*

$$sb - r(z+\kappa) + Ak - (1+r)\psi(k) + \frac{(1+r)(\pi m^h + \tau)}{n} \geq (1+r)\Pi_A^*. \tag{104}$$

Proof. First let L denote the amount of CM goods the bank receive if he follows the proposed allocation. Recall that z is measured in terms of the coming CM goods (after interest received on excess reserves and so on) and is the amount the bank can used to back deposit issuance, and hence, all the IOER promised or other subsidies are included in L. Here we take the stance that the bank pays z units of CM goods in the current CM and receive z in the coming CM. The rest is incorporated in L. Following the proposed allocation, the bank's profit is given by

$$-z - \psi(k) + \varphi b + L + \beta(-b + Ak + z) = \beta\{sb - rz + (1+r)L + Ak - (1+r)\psi(k)\}. \tag{105}$$

Now, the government has to respect fiscal budget constraint. Note that since L is given to the bank outside the proposed allocation for the bank, it has to come from transfers received from the consumers, either through inflation taxes or through banking fees. The total proceeds from those each period is given by $\pi m^h + \tau$, in per consumer terms. Since there are n banks per consumer, we have

$$nL \leq \pi m^h + \tau.$$

Plugging the maximum L into (105), the profit of the bank is then no less than

$$\tilde{\Pi} = \beta\{sb - rz + (1+r)(\pi m^h + \tau)/n + Ak - (1+r)\psi(k)\}.$$

To be incentive compatible for banks to repay κ, it then requires

$$-\kappa + \frac{1}{1-\beta}\tilde{\Pi} \geq \frac{1}{1-\beta}\Pi_A^*, \tag{106}$$

which, after rearranging terms, we obtain (104). $\square$

Since κ and z are completely interchangeable in (104), it is with no loss of generality to set $\kappa = 0$. Finally, we show that Theorem 5.1 remains with the full set of instruments. We combine the necessary conditions:

$$\begin{aligned} -s(b^h + \chi m^h) - (1+r)\tau + \sigma_u[v(y) - p_2] + \sigma[u(q) - d - p_1] - (\chi - 1)m^h &\geq 0, && (107)\\ n[-rz + Ak - (1+r)\psi(k) + sb] + (\iota_m - r)m^h + (1+r)\tau - n(1+r)\Pi_A^* &\geq 0, && (108)\\ sn(\rho Ak + z) \geq snb = sb^h. & && (109) \end{aligned}$$

Now, (107) is the same as (99), (108) follows directly from (104), and (109) follows from pledgeability and market clearing. Now, adding up (107) and (108), we obtain

$$n[-rz + Ak - (1+r)\psi(k)] + \sigma_u[v(y) - p_2] + \sigma[u(q) - d - p_1] - rm^h \geq n(1+r)\Pi_A^*.$$

Now, by (101), $q \leq d + p_1$ and $y \leq p_2$, this then implies (36). Moreover, from (100)-(102) and market clearing, we have

$$q \leq b^d + m^d = b^h + \chi m^h - (\chi - 1)m^d = n(\rho Ak + z) + \chi m^h - (\chi - 1)m^d, \tag{110}$$

$$v(y) - p_2 - (\chi - 1)m^u \geq 0. \tag{111}$$

Now, adding (110) with (111) multiplied by σ_u, we have

$$\sigma_u[v(y) - p_2] + n(\rho Ak + z) + \chi m^h \geq q + (\chi - 1)(\sigma_u m^u + m^d) \geq (\chi - 1)m^h,$$

where the last inequality follows from (102), and this implies (37) since $y \leq p_2$.

B2 Negative externality of informal trades

First note that the presence of externality does not change implementability, and hence Theorem 5.1 is still valid in this case. Moreover, from the proof of Theorem 5.2 we can consider constraints (84)-(86). Thus, with externality given by (41), the social planner's problem is given by

$$\max_{(q,y,k)} \sigma[u(q) - q] + n[Ak - (1 + r)\psi(k)] + \sigma_u[v(y) - (1 + \mathcal{E})y], \tag{112}$$

subject to (84)-(86). The FOC's are still given by (87)-(90), but (89) is changed to

$$\sigma_u[v'(y) - (1 + \mathcal{E})] + \lambda_1\sigma_u[v'(y) - (1 + r)] + \lambda_2\sigma_u v'(y) = 0. \tag{113}$$

Proposition 6.1. *Suppose that $\mathcal{E} \geq r$. Then, optimal $\iota_m > r$, and the optimal IOER is increasing in A.*

Proof. First, one can show that in this case optimal $z > 0$ and hence $\lambda_3 = 0$. This then implies that $\lambda_2 = r\lambda_1$ and that

$$\frac{\sigma[u'(q) - 1]}{r - \sigma[u'(q) - 1]} = \frac{1 + \mathcal{E} - v'(y)}{(1 + r)[v'(y) - 1]}. \tag{114}$$

Let (q^e, y^e, k^e) denote the optimal allocation. By earlier results we know that implementation requires (72) and (73), and hence $s = \sigma[u(q^e) - q^e]/q^e$ and $\iota_m = (1 + s)v(y^e) - 1$. By contradiction, suppose that at optimum $\iota_m \leq r$. For each $s \in (0, \iota_m)$ and , define y_s as the unique solution to $v(y_s) = (1 + \iota_m)y_s/(1 + s)$. Then, by strict concavity of v, $v'(y_s) < (1 + \iota_m)/(1 + s)$. Thus,

$$\frac{1 + \mathcal{E} - v'(y_s)}{(1 + r)[v'(y_s) - 1]} > \frac{(1 + s)\mathcal{E} - (\iota_m - s)}{(1 + \iota_m)(\iota_m - s)} \geq \frac{(1 + s)\mathcal{E} - (r - s)}{(1 + r)(r - s)}.$$

Similarly, define q_s as the unique solution to

$$\sigma[u(q_s) - q_s] = sq_s, \text{ and hence } u'(q_s) < \frac{s + \sigma}{\sigma}.$$

This, together with $\mathcal{E} \geq r$, then implies that

$$\frac{\sigma[u'(q_s) - 1]}{r - \sigma[u'(q_s) - 1]} < \frac{s}{r - s} \leq \frac{(1 + s)\mathcal{E} - (r - s)}{(1 + r)(r - s)},$$

Thus, (114) cannot be satisfied. Therefore, $\iota_m > r$.

Finally, with the same argument as in the proof of Theorem 5.2 it can be shown that optimal q^e increases with A. By (114) it then follows that optimal y^e decreases with A. Thus, optimal IOER increases with A. $\square$