Liquidity, Money Creation and Destruction, and the Returns to Banking*

Ricardo de O. Cavalcanti, Andrés Erosa, and Ted Temzelides†

April 30th, 2004

Abstract

We study a model of money supply in which bank failure may occur as a result of bank illiquidity. In our model, banks issue inside money under fractional reserves, facing the possibility of failure in the event of excess redemptions. Moreover, banks make reserve-management decisions, monitoring the float of their money issue, which affect aggregate liquidity conditions. Numerical examples demonstrate bank failure when returns to banking are low. Central-bank interventions, injecting more funds or making interest payments proportional to holdings of reserves, may improve banks' returns and society's welfare, followed by a reduction in failure rates.

1 Introduction

Monetary theory and macroeconomics have a common history, including a generation of Walrasian models in which notions of liquidity were restricted to measures of a single monetary aggregate. Competitive models are simple to work with, mainly because they abstract

*Preliminary version prepared for the conference *Models of Monetary Economies II*, sponsored by the Federal Reserve Bank of Minneapolis. Comments are welcome.

†The authors are affiliated with the Getulio Vargas Foundation, the University of Toronto, and the University of Pittsburg, respectively.

from the mechanics for "who trades what and with whom." There are, however, important consequences from ignoring how banks and other private institutions create liquidity. In the Kareken and Wallace (1980) volume, a landmark for models of that generation, Tobin (1980) expresses the concern that a model fixing the velocity of money exogenously "...evades all the macroeconomic issues that hinge on the endogenous variation of velocity, questions which involve in turn the menu of money substitutes provided by government or by private agents and intermediaries."

In this paper, we present an alternative approach, incorporating banking into random-matching models of money. We believe that our model provides a useful framework in which one can study the connection between the provision of inside money, the velocity of bank liabilities, and the regulatory environment determining the profitability of the banking industry.

The presentation of our model is proceeded by a review of papers that we find useful, without assuming that the reader is familiar with models of private money. Our goal is to demonstrate that truly private liquidity provision requires a notion of bank liabilities, illiquidity and, to some extent, failure. Otherwise, this liquidity provision can be achieved with *outside money*, without any benefits associated to *inside money*. We model banking in a random-matching environment by assuming that a fraction of the population is no longer anonymous. We assume that a subset of the agents, our banks, can be monitored to some degree and are able to issue substitutes to government currency or outside money. The velocity of these substitutes, the inside money, is endogenous, and depends on the regulation faced by banks. Their behavior concerning money creation is also important and, in turn, it depends on their short-term and long-term payoffs. Our model does allow for excess redemptions of liabilities, with occasional bank failure. As a result, we are able to measure to what extent the private provision of liquidity and the resulting bank illiquidity promote trade and monetary stability.

We compute examples in which banks face a variety of regulations, resulting in different levels of liquidity. Our results might provide some insight on some interesting events in monetary history. Overall, they provide support for the view that a system of privately created liquidity can be self-stabilizing. For example, in one configuration of our model, we find that banks may stop creating money despite its private benefits in the short run. We also find instances, however, in which banks may give up trying to reverse their illiquid position. They instead focus on short-run returns only and eventually fail. Finally, in a case of a liquidity shortage, some infusion of reserves by a central bank may, through general-equilibrium effects on the profitability of private banks, lead to an increase in trade and a reduction in the rate of bank failure.

The paper is organized as follows. We review related models in section 2. There, we also argue that some of the concepts that we shall concentrate on later, such as reserve management and float, can be described in models without explicit money creation. These models make use of a linearity property that facilitates the analysis. We present some of the alternatives allowed by this simplification, but show that it is not compatible with the possibility of bank failure. We then introduce models of inside money, that is, models in which there is private creation of money. We discuss how, when there is perfect monitoring of banks, the reserve management problem is likely to disappear, although other important monetary issues can be studied. One objective of this subsection is to motivate the assumption of imperfect monitoring that we impose later, which not only allows for reserve management and for the float of notes, but also allows for bank failure. We present our model in section 3. There, we also comment on the existence of steady states as well as on properties of bank strategies. In section 4, we use numerical methods to document a variety of equilibrium outcomes. Our conclusion follows. Some proofs appear in the appendix.

2 Literature review

Our model builds on the previous work of Cavalcanti, Erosa, and Temzelides (1999), CET for short. The literature review below is divided between models of banks with outside money, and those with inside money. Kiyotaki and Wright (1989), a building block of our model, allows money to take different forms, such as currency in fixed supply, and/or commodity money created privately. More recent models allow for the use of private fiat money. As we show below, private money can sometimes be substituted by outside money without significant changes. In contrast, in our model banks do create liquidity by providing inside money.

2.1 Outside-money models of banking

We begin by reviewing models of banking in economies where only outside money exists in fixed supply. The meaning of outside money shall become clear as we proceed.

2.1.1 The Wallace and Zhu (2003) model

Wallace and Zhu (2003), WZ for short, present a model of banknote float with divisible production and weak restrictions on money holdings. In this sense, WZ is a generalization of CET. Here we present a simplified version of WZ which ignores these generalizations, but serves as a critical comparison with inside-money models.

Let us consider initially the following standard environment. Time is discrete and the horizon is infinite. There are k perishable goods per date, all of them indivisible. A measure-one continuum of individuals inhabit the economy, and they produce and consume either 0 or 1 unit of a good per date. In section 3, we allow for deaths and births, but for now, let us assume in this section that individuals live forever, and that the common discount factor is $\beta \in (0, 1)$. People specialize in consumption and production. Individuals of type s consume only good s and produce only good $s+1$, modulo k. People cannot commit to future actions, and their histories are to some extent, to be made precise below, private. Individuals meet randomly in pairs once per period, and the probability of meeting with a relevant consumer is the same as meeting with a relevant producer: $\frac{1}{k}$, a fraction independent of s. As is standard, we assume that $k > 2$, so that barter is not possible. We study only steady states with symmetric allocations with respect to s. It is thus helpful to assume symmetric preferences: consumption gives an instantaneous utility u, and production gives a disutility e, with $u > e$. Restrictions on β will be needed in order to demonstrate that a monetary steady state exists, but we postpone that discussion until section 3.

The above environment can be used to derive a role for outside money analytically. That is done by endowing a fraction (of each type) of the population with one unit of fiat money, and by assuming that money is durable but indivisible, and that holdings of money are restricted to either 0 or 1. Here, we consider the following alternative. First, we divide the population of each type into two sets: the banks, of measure B/k, and the non-banks, of measure $(1-B)/k$. We assume that banks can hold money in the form of reserves, r, taking values in the set of integers $\{0, 1, 2, ..., N\}$. The parameters $B \in (0, 1)$ and $N > 0$ may require further restrictions, but they are not important for the discussion that follows. Unlike WZ, we assume for simplicity that nonbanks can only hold 0 or 1 unit of money. More importantly, let us assume as WZ do, that banks can issue notes identified by the name of the issuer, and that banks never meet other banks. That assumption is modified in section 3, when the question of whether a bank issues money to another bank is posed explicitly.

The assumption that notes are indexed by the identity of the issuer gives rise to a reserve-management problem, as a stochastic process resulting from the random trades governs the float of such notes. Formal studies of reserve management date back to Edgeworth (1888) and have been formulated as a partial-equilibrium decision problem of a bank taken in isolation. Informal discussions of how a monetary system can be disciplined by float date back to proponents of the free-banking school and the *Law of Reflux*.[1]

As in CET, banks build reserves by receiving in trade a note issued by another bank. When they do so, the reserve balance of the issuer is reduced by one unit, and the reserve balance of the receiving bank, who deposits the note with a fictitious central bank, is increased by one unit. We preserve the tractability of the model by assuming that all notes are treated the same way by the nonbank population. The next assumption refers to bank failure. We assume initially, as WZ do, that a severe punishment is applied to banks that issue more notes than their reserves (this assumption is relaxed in section 3). The state of a bank at the start of a period is (r, m), where r is total reserves and m is the total number of notes in the hands of nonbanks. The assumption that a severe punishment for failure is enforceable corresponds to imposing $m \leq r$ because a bank in state (r, r) stops issuing notes

[1] See White (1984) as well as Cavalcanti, Erosa and Temzelides (1999) for additional references on the Law of Reflux.

for all r.

Once the number, m, of a bank's notes in circulation is bound by their current r, why should banks care about their float; that is, about how long notes stay in circulation? We assume, as WZ do, that a central bank or government pays interest R on each unit of reserves held per period, with $R > 0$. Hence the total payment of interests is proportional to the length of time during which a note stays in circulation. (In section 3, banks care about float as a result of the possibility of failure, even if $R = 0$, as $r < m$ is allowed). A key property of the WZ model is that it allows the interest to be paid in units of a common good, with a linear utility schedule, without adding much complexity to the model. That linearity also facilitates the reduction of the state space of the economy, as we shall discuss below.[2] We assume that banks receive the interest rate R in a second sub-period after the random meetings take place, and that such transfers are financed by a lump-sum tax not modeled explicitly. We also ignore, in this simplified version, other taxes and clearing fees that WZ consider.

We let p_0^n denote the measure of nonbanks holding 0 notes, divided by k, and let p_1^n denote the measure of nonbanks holding 1 note of some bank, also divided by k. Also we let $p_{r,m}^b$ denote the measure of banks holding r reserves and facing a current liability m in the beginning of the first sub-period, divided by k. A monetary steady state requires these measures to be time invariant, and that nonbanks trade for notes. We also postpone until section 3 a complete description of these stationarity restrictions. For the moment, we let $\phi_{r,m}$ denote the probability that banks choose, as a contingent strategy, to issue a note in exchange for consumption, and let $\gamma_{r,m}$ denote that probability of acquiring a note in exchange for production. Because nonbanks do not receive interest on holdings in the second sub-period, the parameter R is not relevant for them; thus, their values can be simply stated as

[2] In Wallace and Zhu (2003), nonbanks are allowed to hold multiple units of money, and thus a bank may transfer more than one note to a nonbank. The linearity assumption is helpful in that it simplifies the relevant history of a bank's meetings.

$$v_1^n = \beta v_1^n + (p_0^n + \sum_{r,m} p_{r,m}^b \phi_{r,m})[u + \beta(v_0^n - v_1^n)] \tag{1}$$

and

$$v_0^n = \beta v_0^n + (p_1^n + \sum_{r,m} p_{r,m}^b \gamma_{r,m})[-e + \beta(v_1^n - v_0^n)], \tag{2}$$

where v_1^n is the discounted expected utility of starting with one note, and v_0^n is that of starting without a note.

Before presenting the values for banks, we shall state another assumption that simplifies the prediction of floats. As in WZ, we assume now that nonbanks holding a note and meeting with a bank agree to swap their holdings by new notes issued by the bank. By assumption, nonbanks are indifferent about this swap, and banks can use the notes in order to increase reserves and earn interests on float. As a result, the probability that a note held by a nonbank gets retired from circulation, call it π, equals the probability of meeting with a bank, B.[3] In addition, we assume that a bank meets with a nonbank with probability $1 - B$, and meets nobody with probability B. The values for banks thus satisfy

$$\begin{aligned} v_{r,m}^b \quad = \quad & w_{r,m}^0 + p_0^n \phi_{r,m}(u + w_{r,m+1}^1 - w_{r,m}^0) + \\ & p_1^n \gamma_{r,m}(-e + w_{r+1,m}^0 - w_{r+1,m+1}^1) + \\ & (1 - B)kp_1^n(w_{r+1,m+1}^1 - w_{r,m}^0), \end{aligned} \tag{3}$$

where

$$w_{r,m}^j = rR + \beta \sum_i \left(\begin{array}{c} m-j \\ i \end{array} \right) \pi^i (1-\pi)^{m-j-i} v_{r-i,m-i}^b.$$

The last term in the right-hand side of the Bellman equation (3) for $v_{r,m}^b$ corresponds to the increase in float due to the swap of a note. We use the superscript j on the expected utility

[3]In section 3 we do not allow swaps, as swapping would place a small cost or risk to nonbanks. Even if swaps were allowed, however, the probability π would no longer be exogenous since banks are allowed to fail in equilibrium, and a failed bank does not make deposits with the central bank.

from the clearing process $w^j_{r,m}$ to indicate whether the increase in float is due to a note issued in the current period ($j = 1$). The indicator is important as a note issued in the current period cannot be cleared instantaneously since, in the WZ model, banks do not issue notes to other banks. As stated above, a bank meeting with a nonbank holding a note can at least engage in a note swap with payoff $w^1_{r+1,m+1}$. If, in addition, the nonbank is also a potential consumer, an event of probability p^n_1, then the bank chooses, with probability $\gamma_{r,m}$, to add the expected utility $(-e + w^0_{r+1,m} - w^1_{r+1,m+1})$ to his value. Also, as stated above, the stochastic process governing notes is a binomial distribution with parameter π. As discussed, π is exogenous in the present formulation, but the binomial distribution governing the clearing process is the same as that in CET. The term rR is new and corresponds to interest payments on reserves not used or set aside for backing notes in circulation. The formulation is thus one in which such payments affect utility separably, since central banks can transfer units of a common (second sub-period) good to banks.

Thus, this simplified version of the WZ model is a model in which a central bank affects banking decisions through policies of interest payments on reserves. Bank decisions, in turn, affect how much money is put in circulation and, if production were divisible, how much output would be traded, on average, for a unit of money (that is, nominal prices). We now call attention to the fact that the central bank does not allow banks to create money in the WZ model. Related to this is the fact that the central bank is not allowing banks to become illiquid and to risk failure. As a result, the WZ model has an outside-money interpretation. We make that interpretation explicit by presenting a related model where the float of notes is removed. Instead, impersonal outside-money is used in trades. The next proposition discusses the linearity property of $v^b_{r,m}$.

Proposition 1 *Given strategies (ϕ, γ), there exists a unique value function v^b solving the Bellman equation for banks. Moreover, for i such that $v^b_{r+i,m+i}$ is well defined, v^b satisfies the following linearity condition: $v^b_{r+i,m+i} = v^b_{r,m} + Ai$, where A is a positive constant.*

This linearity property can be used to define an alternative banking arrangement, attaining the same allocation as that in WZ. To demonstrate this, we proceed as follows. Instead of using a pair (r, m) in order to denote the state of a bank, let us now consider a regulation or central-bank policy that attaches to each bank as a single state variable, a natural number g. The state g, for "gold," represents net holdings (or wealth), the difference $r - m$ in the original formulation. In order to induce the same behavior as before, we assume that in the second sub-period, R units of the common good are paid per unit of holdings, with gR being paid in total. Next, we demonstrate the equivalence between the WZ economy and an economy with only outside money. We rule out swaps of notes, and assume that when a note is issued in the first sub-period, the balance g is reduced by one unit, and when a note is taken from a nonbank, then the balance g is increased by one unit. It is clear that the proposed simplification treats all money use as if it generated no float, that is, as if all notes issued are redeemed in the same period. We compensate this absence of float with a payment of $\bar{R}$ units of the second sub-period good whenever money is used (that is, whenever g is reduced). By choosing $\bar{R}$ according to the expected forgone interest payments on float,

$$\bar{R} = R \sum_{i=0}^{\infty} \beta^{i+1} \pi (1 - \pi)^i,$$

we are able to generate the same bank decisions. Moreover, in the original formulation, a bank in (r, m) expects the same payment due to floats generated by new swaps as another bank in (r', m'). Thus, the payoff associated to swaps is independent of the current state of a bank, so that the transfers in the second sub-period, corresponding to swaps, can now be omitted without any effect on bank decisions.

We formalize this outside-money economy as follows. With a certain abuse of notation, the Bellman equations for nonbanks remain the same, except that the pair (r, m) is replaced by the scalar g as a bank state. The Bellman equation for banks is now

$$\begin{aligned} v_g^b &= \beta v_g^b + gR + p_0^n \phi_g [\bar{R} + u + \beta (v_{g-1}^b - v_g^b)] + \\ &\quad p_1^n \gamma_g [R - e + \beta (v_{g+1}^b - v_g^b)], \end{aligned}$$

with the understanding that $\phi_0 = 0$.

Thus, in this outside-money version, there are no swaps and no explicit float. The proposition relating this no-float, outside-money economy to the float economy, provided that (p_0^n, p_1^n) does not change, is as follows.

Proposition 2 *If* $(\phi_{r,m}, \gamma_{r,m})$ *is an optimal strategy in the economy with float, then* $(\phi_g, \gamma_g) \equiv (\phi_{g,0}, \gamma_{g,0})$ *is an optimal strategy in the economy without float.*

The main implication of Proposition 2 is that a float economy, with an interest-rate on reserves R, can be duplicated by a no float-economy, with interest-rate on money holdings R, which increases to $R+\bar{R}$ in the last holding period. Thus, the no-float economy loses some of its inside-money interpretation, except that for computing $\bar{R}$, it is necessary to know π, the exogenous redemption probability. The duplication requires setting $p_g^b = \sum_m p_{m+g,m}^b$, since p_g^b must be an invariant distribution implied by (ϕ_g, γ_g), if $p_{r,m}^b$ is an invariant distribution implied by $(\phi_{r,m}, \gamma_{r,m})$. With this construction, (p_0^n, p_1^n) is the same across the two economies.

One of the innovations of the WZ model is to allow a straightforward comparison between a model of private money and one historical episode in the United States before the establishment of the Federal Reserve System, the so-called National Banking Era. National banks were allowed to issue private notes backed by holdings of government bonds. Under the assumption that collateral requirements did disable money creation (an assumption that somehow limits the ability of banks to use other activities to finance bond purchases), the WZ model leads to the conclusion that the system indeed failed to provide an elastic currency regime. Another innovation of the WZ model, in the presence of certain taxes and fees on clearing, is to call attention to the fact that a national bank could rationally choose not to issue a note (not to use all available collateral), whenever the float of that note is expected to be too small. The fact that national banks did not use all available collateral is often called the underissue puzzle. Our version above does not feature underissue, as we did not include taxes and fees in this overview. As we discuss later, underissue may occur in CET if a bank, concerned about the possibility of failure, avoids an m exceeding r. In section 3, we allow meetings among banks, and ask whether bank behavior varies in these meetings, relative to those with nonbanks, just because the former generate no float.

2.1.2 The He, Huang and Wright (2003) model

He, Huang and Wright (2003), HHW for short, propose a model in which gold coins can be lost to theft in random meetings, but notes cannot. They assume that there is an abstract (exogenous) cost in maintaining each unit of note in circulation, which, however, does not apply to gold. The analysis is further extended to allow an endogenous choice between productive and theft activities in random meetings. As a result, a coexistence of gold and notes might arise.

The HHW model thus offers insights about the circulation of alternative media of exchange when their physical attributes differ. Such differences are not the focus of our paper, but there is a section in HHW discussing fractional reserves that merits a comparison with our outside-money version of the WZ model. In the HHW model, banks only meet with nonbanks in what would be the second sub-period, the afternoon, of the WZ model. A bank in the HHW model is in essence a technology for issuing notes, with linear circulation costs, and with an exogenous bound (a "multiplier") on the ratio of notes issued to gold stored. Nonbanks holding a gold coin can obtain a note from the bank after paying a deposit fee of the afternoon good, with a linear utility schedule. Nonbanks without coins can also obtain a note by paying a loan fee of the same good. As a result, there is a limit to the creation of outside money given by the requirement of fractional reserves of gold. In terms of our discussion above, both gold and notes are outside money in the HHW model. Absent their costs of maintaining note circulation (which must be incurred every period and for every note), the random-meetings transactions could be carried by notes only, freeing society from theft.[4]

Our point is that outside-money arrangements do not preclude the possibility of injections of money in the economy. Those injections, done in the afternoon, do not create bank illiquidity in any significant way. They just promote a redistribution of money holdings, in the same way that inflation does in well-known studies of the welfare costs of inflation in random-matching models of money.[5] The HHW model can, however, be used as a benchmark

[4]One way in which this arrangement could potentially be supported, even with circulation costs, would be through the use of lump-sum taxes, which, however, was not considered in their paper.

[5]The study of inflation, under the assumption of a unit upper bound on holdings, is done assuming that monetary injections reach some individuals without money, followed by a lump-sum tax in the form of a probability that a money holder loses his money. We avoid presenting the details here, but one could assume that such a mechanism may be used to implement money injections in versions of the HHW model with

for models that introduce a market in which nonbanks can participate, and where outside money is injected. Next, we discuss an outline of an extension of the WZ model in which such a market is defined, but without explicit references to fractional reserves in gold.

Let us consider again the basic setup of the WZ model presented above, and let us assume that a centralized meeting takes place in each second sub-period, call it afternoon, and that nonbanks without money can also participate in these afternoon meetings. Now both banks and nonbanks can produce and consume afternoon goods, with linear utility in the quantity of the goods. Let us assume further that a central bank intervenes in a market for the creation of money in the following way. Each nonbank in this market has the option to pay ρ, in units of the afternoon good, in exchange for one note. Notes issued to nonbanks in the afternoon must come from a new supply of L reserves issued by the central bank to banks willing to participate in this market. Each bank transfers θ units of the afternoon good per unit of new reserves received, as well as a note that is given to a nonbank. The central bank, thus, collects $(\rho + \theta)L$ with this intervention.

Are there values for (L, R, ρ, θ) consistent with monetary steady states? One contribution of the HHW model is that the linearity of the payoffs for banks and nonbanks may produce a tractable way of determining (R, ρ, θ) endogenously, given measures of outside money and of money injection L. We could perhaps require that in equilibrium $\rho = \beta(v_1^n - v_0^n)$, so that L is demanded, and that θ is given by the linear payoff $\bar{R}$ of the WZ model, so that banks agree to an interior supply of funds. Recall that the central bank is issuing i new reserves and adding i new notes in circulation in exchange for $i\theta$ units of the afternoon goods.

In other words, the HHW structure of two periods, allowing nonbank participation in the second period, can be used to define a market for notes in simple terms. The prices supporting a zero supply of notes in this market ($L = 0$), for instance, are given by $\theta = \bar{R}$ and $\rho = \beta(v_1^n - v_0^n)$, defined by the equilibrium values of the WZ model.

upper bounds.

2.1.3 The Williamson (1999) model

The model described by Williamson (1999), explores other properties of linear payoffs. Like the original HHW model, it allows an aggregation of returns in such a way that the bank sector can be represented by a single bank.

Starting with the HHW model, assume now that nonbanks holding money are also allowed to join the centralized market in the afternoon. In addition, one can assume that the production of afternoon goods requires some investment time, so that the output of the current production is only available for consumption in the afternoon of the next period. We have thus two changes with respect to the HHW model. First, the centralized market can perform an investment function, so that one can explore implications of note issue being backed by investment goods. Williamson (1999) in fact adds more features to this description, such as the possibility that nonbanks invest in autarky, different types of investment projects, asymmetric information about project types, and a stochastic maturity profile for investment projects. Chang (1999), in his discussion of Williamson's paper, argues that the crucial feature in the model allowing notes to circulate, possibly in competition to government money, is the promise of future redemption that is implicitly made by the investment sector.

One question that is not addressed explicitly by Williamson (1999) or Chang (1999) is whether outside money can support the same allocations as banknotes in the model. Since the centralized market can sell the matured goods that resulted from investments completed in the previous period, this market should have the ability to redistribute transfers of money, so that in a steady state all money brought to the market by nonbank consumers is paid out to nonbank producers. Williamson (1999) assumes that banks, or the centralized market in our description, only takes notes issued by the banks themselves and do not accept government currency as payment. This discussion suggests, however, that such an assumption is an inessential part of his model, and that in principle, banks need not care whether a particular means of payment was issued by the government or by another bank, as long as that asset is accepted as reserves by the central bank and can be used to support note issue and other bank activities in the future.

2.2 Inside-money models of banking

We now review models that in contrast to WZ, HHW and Williamson (1999), cannot be described as outside-money economies.

2.2.1 The Cavalcanti and Wallace (1999) model

The main question of the mechanism-design approach of Cavalcanti and Wallace (1999b), CW for short, is whether banking is a robust institution in the context of random-matching models of money. They build on the divisible-goods environment of Shi (1995) and Trejos and Wright (1999). Since these are standard models of outside money, embedding them into a model of banking makes it easier to document the robustness of inside money. However, the main conclusions of CW can be presented in the way that Wright (1999) does, with an indivisible-goods version of CW. We proceed with a brief description in the spirit of Wright (1999). We return to comments about divisible goods and other extensions of CW later.

Unlike the HHW and Williamson (1999) models, bankers in the CW model are individuals facing the same trading opportunities as the nonbank public. Also, there are no sub-periods with common goods as in the WZ structure, so that all bank payoffs come from consumption and production in regular meetings. There are two important points of departure relative to standard random-matching models of money. First, as far as primitives are concerned, being a *banker* is just a label given to individuals that can be monitored. The set of bankers has an exogenous measure $B \in [0, 1]$. When the parameter B is set equal to zero, all individuals are anonymous, producing the standard model as a particular case. Second, given a nontrivial measure of banks, CW ask how the set of allocations resulting when banks are allowed to issue notes (the inside-money regime), compares with the corresponding set when banks cannot create notes (the outside-money regime).

The regime comparison in CW is essentially a study of all equilibrium possibilities. Their approach is to consider the highest punishment allowed by trigger strategies as a device for controlling bank behavior. They find that imposing restrictions on the creation of money does not affect the nature of participation constraints. As a result, they are able to derive a straight comparison between inside and outside money: any equilibrium that uses only outside money can be duplicated as an equilibrium that uses only inside money. Moreover, the allocation that maximizes an average-welfare criteria over the set of inside-money allocations, which is an equilibrium (satisfying participation constraints) for some parameters, cannot be achieved by using outside money.

We use the following notation in order to state these findings formally. Nonbanks carry either 0 or 1 unit of notes as before. Money is distributed symmetrically across the k specialization types according to p_0^n and p_1^n, with $p_0^n + p_1^n = (1 - B)/k$. Depending on the regime, the state j of a bank has a different interpretation, but banks of all types are in general distributed symmetrically across states according to p_j^b, with $\sum_j p_j^b = B/k$. Let us now consider the welfare criterion Z,

$$Z = k^2 \sum_{i,j,i',j'} p_j^i p_{j'}^{i'} (u_{jj'}^{ii'} - y_{jj'}^{ii'}),$$

where $i \in \{b, n\}$ denotes the type bank/nonbank and j the state of an individual, with primes used to distinguish the consumer, $y_{jj'}^{ii'} \in \{0, e\}$ indicating whether the (i, j) potential producer actually produces for the (i', j') consumer, $u_{jj'}^{ii'} = u$ if $y_{jj'}^{ii'} = e$ and $u_{jj'}^{ii'} = 0$ otherwise. It can be shown that the average discounted utility in this simple economy is proportional to Z since the product $p_j^i p_{j'}^{i'}$ measures the frequency of the meeting between the (i, j) producer and the (i', j') consumer, and $u_{jj'}^{ii'} - y_{jj'}^{ii'}$ measures the net payoff in this meeting. Hence the average payoff must be proportional to Z.

As usual, nonbanks switch states when they acquire and spend money. The way in which banks switch states depends on the way they are regulated; in particular, it depends on whether they can create money. CW define the outside-money regime as a regulation forcing banks to only use outside money, like in the WZ model. If banks can only hold 0 or 1 unit of outside money, like the other agents, then the state j for banks can be used as an index to money holdings. With the objective of maximizing Z, it is a good idea to have banks producing as often as possible. A desirable allocation has banks producing for nonbanks even when they cannot pay with money. As Wright (1999) points out, there is also no need for banks to ask for payment from the non-banks that do have money. Although the bank could use the money to buy goods in the future, there is no net gain in terms of Z

because the nonbank would then leave the meeting without this purchasing power. Hence, a desirable outside-money allocation has the banks producing gifts to nonbanks with and without money.[6] Imposing this outcome, the value functions for nonbanks can be written as

$$v_1^n = \beta v_1^n + \sum_j p_j^b u + p_0^n[u + \beta(v_0^n - v_1^n)]$$

and

$$v_0^n = \beta v_0^n + \sum_j p_j^b u + p_1^n[-e + \beta(v_1^n - v_0^n)].$$

The participation constraint for the nonbank consumer is $u + \beta v_0^n \geq v_1^n$, and that for the nonbank producer is $-e + \beta v_1^n \geq v_0^n$, but it is easy to show that the latter implies the former.

As indicated in the Bellman equations for nonbanks, all banks produce for nonbanks in single-coincidence meetings and, without loss of generality, they do not request money payments under the outside-money regime. Thus the state j of a bank is irrelevant for nonbanks. Likewise, it is desirable to have banks producing for other banks whenever possible, and since banks can be monitored and punished with autarky, there is no need to use money in these transactions either. Writing now the Bellman equation for banks as

$$v_j^b = \beta v_j^b + \sum_i p_i^b(u - e) + \sum_i p_i^n(-e),$$

their participation constraint is $-e + \beta v_j^b \geq 0$, due to the following: after producing and incurring disutility e, the expected value for a bank is βv_j^b, while the payoff from deviating is 0, because all individuals meeting with this bank can be instructed to never produce for him once a deviation is recorded.

Let us suppose now that banks are allowed to issue notes. The duplication result of CW states that one can reproduce any allocation that is achieved with the use of outside money by using note creation and destruction instead. It is important to remark that this result applies to any outside-money allocation, even to the one in which banks are instructed to accept payments (a possibility not allowed in the value functions above for simplicity). In

[6] This result, however, does not hold with divisible goods when participation constraints bind.

order to state a limited version of the duplication result (which CW call the strict-subset result), we proceed as follows. First we fix the outside-money allocation, say by choosing one in which gold coins are used by everybody, and construct an allocation without gold but with the creation and destruction of bank notes. If any bank is given gold in the outside-money allocation, then a state $j = 1$ is now assigned to him, and when that gold is spent by the bank in the outside-money allocation, this bank is now allowed to print his own money and exchange it with a consumer. After that, his state is adjusted to $j = 0$. A bank in $j = 0$ is prohibited from printing money and only moves to $j = 1$ in meetings where he receives a note from a nonbank. It can easily be verified that the distribution of notes is the same now as the distribution of gold in the outside-money allocation.

It also follows that this duplication is accomplished without reference to the welfare criterion Z. Simply put, any outside-money allocation can be duplicated because the participation constraint $-e + \beta v_j^b \geq 0$ stays the same across both regimes as long as the law of motion of the state j is the same as that of money holdings with outside money.

The duplication result may seem obvious since, after all, it is just a statement that monitoring and record keeping can substitute for outside-money use when banks are perfectly monitored. However, it demonstrates that the differences between outside and inside money do not pertain to the physical characteristics that money might exhibit, or even to the aggregate quantity of money, but, rather, to the possibilities of creation and destruction that a simple measure of monetary aggregates may not detect. In order to illustrate this last point, let us consider now an inside-money allocation for the above economy, in which banks are told to request payments from nonbank (money holding) consumers. They are also allowed to print notes and use them to buy goods from nonbank producers. The value functions for nonbanks now change to

$$\bar{v}_1^n = \beta \bar{v}_1^n + (\frac{B}{k} + p_0^n)[u + \beta(\bar{v}_0^n - \bar{v}_1^n)]$$

and

$$\bar{v}_0^n = \beta \bar{v}_0^n + \frac{B}{k} u + (\frac{B}{k} + p_1^n)[-e + \beta(\bar{v}_1^n - \bar{v}_0^n)].$$

The values $\bar{v}_i^n$ now represent expected discounted utilities under this inside-money regulation. Notice that we did not assign a state to bankers in these equations because we shall treat all banks the same way, independently of their histories. As before, banks continue to give gifts when they meet nonbank consumers without money. When they meet nonbank producers without money, they issue a note to them in exchange for production. The value for a bank is $\bar{v}^b$, where

$$\bar{v}^b = \beta \bar{v}^b + \frac{B}{k}(u-e) + p_0^n(u-e) + p_1^n(-e).$$

Participation constraints for banks and nonbanks do not change. We then have the following.

Proposition 3 *Average welfare with inside money is greater than that with outside money by at least* $B\frac{1-B}{2}(u-e)$.

The welfare differential stated in Proposition 3 is only applicable when β is sufficiently high, so that participation constraints do not bind. The highest welfare attained by outside money takes place when half of the nonbanks do not hold money, which is the same distribution achieved by the inside-money allocation discussed above. Hence, although the quantity of money can be the same in both allocations, inside money allows a higher welfare by the difference stated in the proposition because banks are now allowed to consume in meetings with nonbanks without money, and the frequency of these meetings is given by $B\frac{1-B}{2}$. If outside money is not distributed optimally, then the welfare differential is even higher. If $B = 0$, then all individuals are nonbanks and the inside-money allocation collapses to one that can be duplicated with outside money chosen optimally. As pointed out by Wright (1999), the CW approach allows the distribution of money to be chosen as part of optimal allocations and, as a result, it establishes the robustness of inside money.

2.2.2 Refinements of the CW model

There are refinements of the CW model that add to the understanding of the properties of inside money. We shall give a brief description of these developments before we discuss the approach of CET, which contains related findings, and which is the basis of our experiments relating liquidity to the profitability of the banking sector.

Cavalcanti and Wallace (1999a) contains an extension of the CW model in which individuals face idiosyncratic productivity shocks that are private information. Although they allow for divisible production, they characterize the optimum for parameters in which the nonbank participation constraint does not bind. They show that, when the size of the banking sector is small and the probability of an adverse shock is also small, the optimum is implemented by assigning banks a state variable assuming two values. Bank consumers in the high state are allowed to consume from nonbanks. Bank producers in the high state who announce that they received an adverse productivity shock are transferred to the low state. Bank consumers in the low state receive a low consumption from nonbanks, and are only transferred back to the high state after demonstrating, by showing their production, that they received the high productivity shock. The authors thus provide an example of inside money in which banks are assigned states that are not governed by their histories of creation and destruction of notes, but, rather, by a system that monitors their production announcements directly.

Mills (2001) extends the CW model in a different direction. He assumes that banks may, with some probability, engage in anonymous trades. He shows that optimal allocations may require the use of both inside and outside money. In Mills's model, outside money can become necessary because outside money holdings can be used as evidence that banks produced in anonymous meetings. One crucial assumption in Mills's model, however, is that inside money is uniform and not distinguished by the identity of the issuer. Otherwise, notes issued by other banks, excluding a bank's own notes, can be used as evidence of production as in the models of WZ and CET. Mills's model is nevertheless a nice illustration of how certain limitations on the set of available mechanisms can produce a co-existence of inside and outside money.

The review of the CW model presented by Wright (1999) may give the impression that the only welfare gains allowed by inside money are due to an increase in bank consumption. More recently, Cavalcanti (2004) considers a matching model in which the meetings between banks and nonbanks are all unproductive, except that fiat and productive assets can change hands. He shows that banks, being perfectly monitored as in CW, can be regulated so as to use credit arrangements among themselves, which dispenses with the use of money, but which leads to an efficient reallocation of a productive asset termed capital. The author first compares this credit arrangement with one in which outside money is used by nonbanks who are prohibited from trading with banks. He shows that banks are able to generate a more efficient allocation of capital than nonbanks can via the use of (outside) money. The author then allows nonbanks to open "deposit" accounts in meetings with banks through a system that lets banks to compare identities with passwords. With this system, nonbanks remain anonymous in meetings with other nonbanks, but their histories of deposits and withdrawals with banks is recorded and made available to all other banks. Under some conditions, which include a degree of capital scarcity, it is shown that, at the optimum, inside money is issued to nonbanks in exchange for capital, which is in the future intermediated to other nonbanks.

Cavalcanti's model points out some new implications of the property that inside money can be created with less restrictions imposed by past histories, in comparison to outside money. This additional liquidity can facilitate the allocation of other resources, such as capital, and support welfare gains that go beyond changes related to bank consumption. One possible avenue for future research is to study how interest payments in the form of inside money, or, possibly, in the form of common goods like in the WZ model, may be used to improve the allocation of intermediation services.

Cavalcanti and Forno (2003a) perform another study of the role of regulation in the CW model. They start with the optimal inside-money allocation for a high β such that nonbank participation constraints are weakly binding, and then reduce β continuously in order to investigate how the optimum changes. When β is high, then nonzero production levels, those maximizing the welfare criteria Z, equalize marginal utilities, $u'(y^{ii'}_{jj'})$, to one, so that $u^{ii'}_{jj'}$ is equal to a constant u that is independent of the state. Thus, for high β, the role of inside money identified by Wright (1999) emerges. Cavalcanti and Forno show, however, that as β is reduced and the nonbank participation constraint starts to bind before that of banks, then the optimum features no changes in the distribution of money, but it involves banks redeeming notes with more production than what they receive when issuing them; in other words, banks start paying interest on notes. The intuition behind this result is that by offering a higher return on note redemption, banks increase the ex-ante return on money, alleviating the nonbank participation constraints. Their result supports an alternative to the Friedman rule because other factors, in addition to the quantity of money, determine the return of inside money in the CW model. Using numerical methods, Cavalcanti and Forno (2003a) extend their model by adding aggregate preference shocks to the CW environment. They find that not only do the interest rates differ from contractions to expansions due to differences in the severity of participation constraints, but welfare also improves by allowing interest rates to be non-stationary. In particular, optimal interest rates may be autocorrelated even when shocks are *iid*.

Finally, another extension of the CW model provides a closer connection between bank profits and the provision of liquidity. Cavalcanti and Forno (2003b) study bank competition by allowing banks to choose one out of two possible networks. Banks in a network issue notes of the "same color," but that color is distinct from that of the other network. They demonstrate the existence of equilibria in which networks feature two opposing externalities. The credit externality refers to the property that as the mass of banks in one network increases, their members can trade among themselves using future promises. These promises become more valuable as the network size increases due to the higher frequency of meetings inside the network. The money externality refers to the property that with the increase in network size, more notes of that network find their way into circulation, increasing the monetary liability of each member of that network. The authors find that when B, the measure of the bank sector, is small, there exists an ex-ante equilibrium in network choice that is stable in a particular sense. That stability is lost as B increases and credit dominates the money-externality negative effects. In that case, the equilibrium features a monopoly in note issue, resembling the planner solution of CW.

The notion that monetary liabilities can have direct effects on bank behavior and reinforced effects on the overall provision of liquidity in the economy is a main topic of discussion in the CET model, which we discuss next.

2.2.3 The Cavalcanti, Erosa, and Temzelides (1999) model

The first paper to introduce inside money in medium-of-exchange models was, to our knowledge, Cavalcanti, Erosa, and Temzelides (1999), CET for short. Their model has two distinguished features. First, they modelled the creation of circulating bank liabilities without costs to money creation. In other words, the subject of CET was fiat money created privately. In this regard, the paper attempted to address the question posed by Fisher (1986) of what he called the dynamic inconsistency of private money. In a monetary equilibrium, one in which money has value, how could private agents create costless money without promoting monetary instability? Fisher argued that promises of conservative creation would be reneged, and that existing theory would give more support to Friedman's views, which defended the government provision of money, rather than that of Hayek, who favored currency competition. The arguments in Fisher (1986) are provocative, but also seem too informal. Without a formal model in which private money can be created systematically, it is hard to provide insights about historical episodes in which financial stability was achieved, or current and future instances in which means of payments become increasingly private. The lack of models of private money was eloquently documented in a survey by King (1983).

CET offer an alternative to the dynamic inconsistency pointed out by Fisher (1986). They argue that some degree of monitoring can support a stable banking system. The monitoring in CET was related to record-keeping of reserves. The punishment for negative reserves was assumed to be bank dissolution, with the failing bank becoming anonymous but not isolated from nonbank life.[7] Under decentralized random meetings, and without access to centralized markets, banks could not use arbitrage opportunities to generate unbounded

[7] This specification is in some agreement to the *free banking era* in the United States (1837-1865). During that period, banks were penalized for failing to convert their notes into "specie" upon demand. In fact, at least in some states, if the bank failed to redeem (notes plus some penalty fee) within a grace period, the bank was forced to close down. Interestingly, while banks were by law obliged to redeem their notes for specie, notes issued by other banks were typically accepted as a substitute. Under typical (but certainly not all) circumstances, there was little or no discount between private notes issued by banks in good standing and specie. On the other hand, notes issued by banks that were likely to fail traded at great discount. See Gorton (1996) for a detailed discussion.

profits instantaneously. A strategy of always issuing money, without building reserves, could only add utility at discrete increments over time. Thus, the punishment of bank dissolution, with a small but positive probability, could render this strategy too risky, since a dissolved bank would lose all the prospective profit opportunities of a conservative bank. Additional explanations of why steady states with creation of inside money by banks may not only exist but improve welfare were later developed in versions of the CW model. As seen above, these versions choose the bank's payoff and punishment optimally, among the available alternatives that respect individual rationality.

A second feature of CET was the modeling of reserve management by banks in an environment with float, fractional reserves and possible bank failure. They used the simplifying assumption that a bank consumer could not distinguish whether a producer was a bank or a nonbank. In the next section, we modify this assumption, by allowing banks with unfavorable reserves to avoid note issue in meetings with other banks. We use numerical methods and phase diagrams to illustrate bank behavior in the (r, m) space. In the computed examples, some banks with $r < m$ do not issue to other banks, but issue to nonbanks in order to maximize the payoff from float. Thus, float payoffs need not result from interest payments on reserves, but can also result from the attempt to reach a balance between a fractional-reserve policy, with good short-run consumption opportunities, and a tolerable risk of bank failure.[8] Banks in CET are allowed to issue as many notes as they want and are thus allowed to become illiquid. One of the main findings in CET is that the float of notes is also

[8] On the one hand, this finding is in some sense contrary to Friedman's pessimistic view on free banking, as expressed in his "Program for Monetary Stability." He argues that in situations where individuals engage in transactions with others that are far removed in "space and acquaintance, and [in which] a long period may elapse between the issue of a promise and the demand for its fulfillment," it is likely that profit incentives will result in unsound practices, which will result in the instability of monetary trades. On the other hand, it is clear that there are very diverse experiences regarding the performance of monetary systems with an explicit use of private money. The Scottish banking system, for example, is widely recognized as being successful, while the American experience of the nineteen century is mixed. Even within the same period, and the same country, the experience of Michigan, Illinois, and Wisconsin during the free banking era is considered to be a failure, while the experience in New York is considered to be a success. See King (1983) and Rolnick and Weber (1983), (1984) for detailed discussions.

governed by bank policies, since banks are endogenously choosing how many notes to issue and how many notes to destroy (by building reserves). Thus, general equilibrium effects link banking behavior to float averages and back to bank behavior. CET did not explore these possibilities, nor did they investigate how the parameters of the model, including policies injecting money into private-banks' reserves, would interact with such general equilibrium effects. We explore some of these interactions in section 4.

3 Our model

Our model contains elements of both CET and WZ. With exception of the demographics, the goods-structure is that of our review of WZ. We include in each date both the random-meeting goods (the number of types is $k > 2$), and the general good, used to pay interest on reserves at rate R, and to collect the taxes financing these payments. The utility of consuming general goods (earning interest) and the disutility of producing general goods (paying taxes) are linear (with slope one). Hence, with lump-sum taxation, it suffices to subtract from the equilibrium value functions, when $R > 0$, a constant that equals total interest paid to the bank sector. For simplicity of the presentation, we ignore this constant (the tax rate) from our Bellman equations, although we make the necessary adjustments when computing welfare measures.

As before, we limit nonbank holdings to $i \in \{0, 1\}$. The state of a bank is (r, m). Unlike the WZ model, the level of reserves of a bank, r, can be less than the number of notes in circulation, m. There exists a central bank that shuts down banks that experience a number of redemptions, in the end of a period, exceeding reserves. The old notes of a failed bank in circulation remain valid and are accepted in trade, because the central bank "honors" these liabilities by accepting them, as deposits of reserves from any other bank, at any future date. A failed bank is, however, prohibited from issuing new notes. Hence all notes in circulation are valid notes, and they never pose a risk to nonbanks. In order to produce the simplifying result that banks always deposit notes with the central bank, even if $R = 0$, we assume that a failed bank exits the banking industry with a valid note and continues to trade as a nonbank. Hence, depositing notes is a dominant strategy for active banks, even in the absence of interest payments on reserves. We do not need to distinguish failed banks from the original nonbanks.

In order to accommodate the possibility of failure within the steady-state analysis, we assume that there is constant inflow of new individuals in the population, both banks and nonbanks, and there is a constant outflow. That is accomplished by the simple assumption that, with probability $\delta \in (0,1)$, an individual dies and is replaced by another individual with the same type. The notes of a bank that dies in circulation remain valid for deposits with the central bank. A note held by a nonbank that dies is cleared by the central bank, so that δ introduces a lower bound on the probability of redemption π. As said above, the probability π is no longer a parameter. We do not allow swaps of notes in meetings and, since the measure of active banks is endogenous, the reserve-management behavior of banks determines π. The effective discount factor, after death rates are taken into account, is $\beta = (1-\delta)\tilde{\beta} \in (0,1)$, where $\tilde{\beta}$ is the original discount rate. Unlike WZ, we assume that banks meet with other banks and, unlike CET, we assume that banks know when their trade partner is a bank (who deposits notes immediately).

We denote the nonbank expected discounted utility, before taxes, by v_i^n, and that for a bank by $v_{r,m}^b$ ($w_{r,m}^j$ just before redemptions are computed). Their present value is zero when dead. The sequence of events for within a period is displayed in Figure 1.

$t \rightarrow$ possibility of death $\rightarrow v_i^n$, $v_{r,m}^b \rightarrow$ meetings $\rightarrow$ trade and
deposits $\rightarrow w_{r',m'}^j \rightarrow$ reserves updated and interest paid $\quad t+1$

Figure 1: Sequence of Events

We let p_i^n denote the measure of nonbanks holding $i \in \{0,1\}$ unit of money and $p_{r,m}^b$ denote the measure of banks in state (r,m) before trade meetings. Due to discounting, spending money is a dominant strategy for nonbank consumers. Thus, the only relevant decision for nonbanks refers to the acceptance of money in exchange for production. Clearly, a nonbank producer in state $i=0$ engages in trade whenever $\beta(v_1^n - v_0^n) > e$. We shall restrict parameters so that this is indeed the case in monetary steady states.

When banks in state (r, m) are in a meeting where they can produce, they improve their reserves position by accepting money in exchange for production, with probability $\gamma_{r,m} \in [0, 1]$. Notice that while nonbanks always agree to exchange money for consumption, that decision is nontrivial for banks. We shall see that a concern about building reserves can induce banks to behave differently. When deciding whether to issue a note, in exchange for consumption, their behavior may depend on their reserve positions and also on whether the recipient of the note is a bank or a nonbank. We let $\phi^n_{r,m}$ and $\phi^b_{r,m}$ denote the probabilities that a bank in state (r, m) issues a note to a nonbank and to a bank, respectively. With this notation, the value functions for nonbanks are given by equations (1) and (2) of the WZ model, with $\phi_{r,m}$ replaced by $\phi^n_{r,m}$.

If a bank issues a new note to a recipient that is also a bank, the fact that the new note is deposited and does not enter circulation is marked by adding to the bank state, before redemptions are computed, the indicator $j = 0$. If the new note does enter circulation, because it has been issued to a nonbank, the indicator instead is $j = 1$. After reserve balances are finally updated, banks infer the quantity of notes that remain in circulation. Taking into account that banks that reach a negative reserve balance exit the industry with one note, one finds that the current $w^j_{r,m}$ depends on next period values according to

$$w^j_{r,m} = rR + \beta \sum_{i \le r} \tbinom{m-j}{i} \pi^i (1-\pi)^{m-i-j} v^b_{r-i,m-i} + \beta \sum_{i > r} \tbinom{m-j}{i} \pi^i (1-\pi)^{m-i-j} v^n_1, \qquad (4)$$

with the understanding that the sum is taken with respect to i, with values ranging from 0 to $m - j$ possible redemptions. Considering now that the number of banks willing to create money in meetings with other banks is given by the inner product $\phi^b \cdot p^b$, it follows that the fraction of the population willing to purchase from a bank is $\phi^b \cdot p^b + p^n_1$. Thus, a bank's value is given by

$$\begin{aligned} v^b_{r,m} \quad = \quad & w^0_{r,m} + (\phi^b \cdot p^b + p^n_1) \max_{\gamma_{r,m}} \gamma_{r,m}(-e + w^0_{r+1,m} - w^0_{r,m}) + \\ & p^n_0 \max_{\phi^n_{r,m}} \phi^n_{r,m}(u + w^1_{r,m+1} - w^0_{r,m}) + \\ & \gamma \cdot p^b \max_{\phi^b_{r,n}} \phi^b_{r,m}(u + w^0_{r-1,m} - w^0_{r,m}), \end{aligned} \qquad (5)$$

where, as before, u and e are the utility from consumption and the disutility from production, respectively, with $u > e$.

We now turn to the description of the supply of money. There are two sources of money injection. The first source is exogenous. We assume that every period, a fraction μ_n of the newborns are nonbanks receiving 1 monetary unit. Hence, $\mu_n\delta$ equals the inflow measure of outside assets. The second source is endogenous. We also assume that a fraction μ_b of the newborns, where $\mu_b < 1 - \mu_n$, are each endowed with a banking technology. The stationary measure of potential banks is μ_b, and that of nonbanks is $1 - \mu_b$. If there is no failure in equilibrium, then the parameter μ_b would be the counterpart of the parameter B of the models reviewed above. Besides the newborns starting with outside money or with the banking technology, there exists a fraction of nonbank newborns, $1 - \mu_n - \mu_b$, starting with zero monetary units. We shall restrict attention to outcomes in which private money is treated as a perfect substitute to outside money in all trades, and we use our notation accordingly. In particular, there is no need to decompose the state of a nonbank, i, into holdings of either outside or inside money.

The final aspect of bank regulation is the assignment of initial states to newborn banks. Let us suppose that all newborn banks are assigned $(r, m) = (\bar{r}, 0)$ as an initial state, and let us consider momentarily an arbitrary cut-off reserve level r^*, for determining failure. If $R = 0$, it is clear that the same banking behavior should follow from assigning $(\bar{r} - r^*, 0)$ as an initial state and changing the cut-off level to zero. Hence, with respect to policies that keep $R = 0$, we can, without loss of generality, fix the lower bound for reserves as $r^* = 0$ and consider different choices of initial balances $\bar{r}$. We do so assuming that the state of a newborn bank is $(\bar{r}, 0)$. The distributions (p^n, p^b) must remain invariant, given the interest-rate policy, the money-injection parameters, the bank strategies, and the fact that nonbanks accept money in trade. The restrictions imposed by stationarity are described in the Appendix, by means of an operator mapping a current distribution into one for the next period. An invariant distribution is a fixed point of that operator. Among the statistics of interest, which the fixed point allows to be computed, are the probability of redemption, π, and the measure of bank failure per period, z.

In summary, the money supply is endogenous and depends on banking regulation and monetary policy. Monetary policy is characterized in the model by μ_n, describing the distribution of outside money given to nonbank newborns, by $(\mu_b, \bar{r})$, describing the entry rate and reserve position of new banks, and by R, describing the interest rate on reserves. The next subsection briefly discusses the existence of a stationary symmetric equilibrium exhibiting certain properties.

3.1 Existence of steady states

In a steady state, banks and nonbanks make their consumption and savings decisions taking as given the policy parameters and the variables p^n, p^b and π, as well as the decision of others. We now discuss how these variables are determined in equilibrium. We let T_{γ,ϕ^n,ϕ^b} denote the continuous mapping, implied by any given list of strategies (γ, ϕ^n, ϕ^b), which maps (p^n, p^b) into the next period's distribution of individuals across states (defined in the Appendix). We define steady states as follows.

Definition 4 *A monetary steady-state equilibrium is an array $(v^n, v^b, w, \gamma, \phi^n, \phi^b, p^n, p^b, \pi)$ satisfying the Bellman equations (1)-(2), with $\phi = \phi^n$, as well as (4)–(5), and such that $\beta(v_1^n - v_0^n) > e$; (γ, ϕ^n, ϕ^b) attains the maximum in (5); (p^n, p^b) is a fixed point of T_{γ,ϕ^n,ϕ^b}; and $\pi = \delta + (1-\delta)\gamma \cdot p^b$.*

Regarding the last equilibrium condition, we notice the following. Inside money held by nonbanks is redeemed upon death. Its is also redeemed when nonbanks survive but trade in meetings with bank producers, an event occurring with probability $(1-\delta)\gamma \cdot p^b$. As a result, $\pi = \delta + (1-\delta)\gamma \cdot p^b$.

In every period, there is a constant inflow of newborn nonbanks without money, of measure $\delta(1 - \mu_n - \mu_b) > 0$. This fact can be used to derive a condition (see Lemma 1 below) relating β, u/e, k and $1 - \mu_b - \mu_n$, so that $\beta(v_1^n - v_0^n) > e$ holds for any fixed point (p^n, p^b). This condition suffices to produce a degree of scarcity of money, relative to β and to u/e, so that money has value independently of bank behavior, provided that a stationary distribution of banks exists. We go beyond this simple argument and show that, as β is chosen sufficiently high, bank behavior becomes more conservative and oriented to building reserves, up to the point that bank failure is essentially eliminated. Without loss of generality, we proceed by fixing $R = 0$ (positive values of R tend to give money intrinsic value, facilitating the existence of steady states). In this case, however, given a fixed β, banks with sufficiently high reserves choose to issue notes when $m = r$, thus becoming illiquid. That is so because, for m sufficiently high, the probability that all m notes are cleared at the same time becomes arbitrarily low.

More formally, our argument on the existence of a monetary equilibrium with banking proceeds by following similar steps at those in CET. We restrict attention to the case in which $R = 0$, and every newborn bank starts with $\bar{r} = 0$ reserves (although the numerical experiments use different specifications).We first restrict nonbanks to accept money (as stated in the Bellman equations for nonbanks). Next we restrict banks to back note issue by 100% reserves on a range of the state space indexed by a constant S (banks in this range are also restricted to accept money). The restrictions for banks read $\gamma_{r,r} = 1$ and $\phi^b_{r,r} = \phi^n_{r,r} = 0$ for $r \leq S$. Non-restricted banks choose their strategies optimally. We then construct a correspondence, ψ, producing new steady-state candidates from any given array $(v^n, v^b, w, \gamma, \phi^n, \phi^b, p^n, p^b, \pi)$, customarily, as follows. We determine new lists for (γ, ϕ^n, ϕ^b) according to the above restrictions for $r \leq S$, and as the correspondence of strategies that attains the maximum in the right-hand side of Bellman equations, given the old list $(v^n, v^b, w, \gamma, \phi^n, \phi^b, p^n, p^b, \pi)$, for the unrestricted states. The restrictions and maximizations also give a new triple (v^n, v^b, w) as the left-hand side. Next, the new measures over states are given by T_{γ,ϕ^n,ϕ^b} applied on (p^n, p^b), and π is computed according to the formula $\pi = \delta + (1 - \delta)\gamma \cdot p^b$. The existence of a restricted steady-state equilibrium then follows since ψ satisfies the conditions of Kakutani's fixed-point theorem.[9]

Next, we argue that the restrictions are nonbinding when β is sufficiently large. To an individual nonbank, the fact that other nonbanks without money, as well as newborn banks, are willing to accept money with probability one implies a lower bound on $(p^n_0 + p^b \cdot \gamma)$. Thus,

[9] We assume an upper bound (N, N) on the state space for banks, so that this space becomes compact. Because $\delta > 0$, the measure of banks who survive until they reach a state outside this upper bound can be made arbitrarily small by the choice of a sufficiently large N. Hence, this compactness restriction is innocuous.

if β is sufficiently high, the restriction $\beta(v_1^n - v_0^n) > e$ is nonbinding (see also Lemma 1).

For β sufficiently large, the restrictions for banks are also nonbinding. Notice that if $\phi_{r,r}^b = 0$ and $\phi_{r,r}^n = 0$ are nonbinding for $r \leq S$, then $\gamma_{r,r} = 1$ is also nonbinding for $r \leq S$; otherwise, banks would be choosing a form of autarky after reaching a level of reserves $r \leq S$. We now show that $\phi_{r,r}^b = \phi_{r,r}^n = 0$ is indeed nonbinding for $r \leq S$. It suffices to show that $v_{0,0}^b - v_1^n$ can be made arbitrarily large, by picking β sufficiently close to one in a restricted steady state. If $\phi_{r,r}^b = 1$, a bank faces, with probability 1, the possibility of losing $v_{0,0}^b - v_1^n$. The relevant case, thus, concerns the case where a bank consumer meets a nonbank producer.

If $\phi_{r,r} = 1$, a bank faces, with positive probability, the possibility of losing $v_{0,0}^b - v_1^n$ after one period in the event of excess redemptions. From the Bellman equation for banks, we know that $\phi_{r,r}^n = 0$ is implied if $u + w_{r,r+1}^1 < w_{r,r}^0$ holds or, equivalently, if $u < w_{r,r}^0 - w_{r,r+1}^0$. Now, $w_{r,r}^0 - w_{r,r+1}^0$ is bounded below by $\beta\pi^r(1-q_b)(v_{0,0}^b - v_1^n)$, where π^r equals the probability that all notes are redeemed and $(1-q_b)$ represents the probability that the bank cannot increase reserves in the next period $(q^b = p_1^n + \phi^b \cdot p^b)$. Thus, $\phi_{r,r}^b = \phi_{r,r}^n = 0$ if $u < \beta\ \pi^r(1-q_b)(v_{0,0}^b - v_1^n)$ for all $r \leq S$. Since $\pi^r \geq \delta^r > 0$ and $1 - q_b > \frac{(k-1)}{k}(1-\mu_n)\delta > 0$, the conclusion follows if $v_{0,0}^b - v_1^n$ is sufficiently large. The same argument as in CET (p. 941) can now be applied to show that $v_{0,0}^b - v_1^n \to \infty$ as $\beta \to 1$. This completes the existence argument.

3.2 Derivation of key strategies

In this section, we describe optimal strategies of any given monetary steady-state equilibrium with $R = 0$. We first establish some properties of the value functions and of the banks' strategies. These properties are of independent interest but they are also used in guiding the computational experiments that we perform in the next section. We start with a sequence of preliminary Lemmas. The proofs appear in the Appendix. The first Lemma gives a sufficient condition for nonbanks to accept money (outside or private) in exchange for production. The proof is similar to that in CET and is thus omitted.

Lemma 1. *If* $\frac{1-\mu_n-\mu_b}{k} \geq \left(\frac{1}{\beta} - 1\right)\frac{1}{\frac{u}{e}-1}$, *then* $\beta(v_1^n - v_0^n) \geq e$.

Note that the condition of Lemma 1 will tend to be satisfied if the discount factor is large and if the fraction of newborns that are banks is small. If the fraction of banks is too high, a steady-state equilibrium where money is valued might not exist. It is also intuitive that nonbanks accept money when the ratio u/e is sufficiently high.

The next Lemma asserts that the value function of a bank is increasing in the amount of reserves and decreasing in the amount of liabilities in circulation.

Lemma 2. *The value function $v^b_{r,m}$ is weakly increasing in r and weakly decreasing in m.*

Our next lemma says that the value of a bank is always greater than the value of a nonbank holding one unit of money which, in turn, is greater than the value of a nonbank with no money holdings. Banks can do everything that nonbanks can and, in addition, they can build reserves and "borrow" by costlessly issuing new money. This additional value of being a bank, which partly relies on assumptions that guarantee limited entry into the banking sector, plays an important role in the numerical results, in which we study the effects of central-bank policies on note issue. In particular, policies injecting outside money (including changes on initial reserves) will affect bank returns, and, thus, the behavior of banks regarding issue.

Lemma 3. *For all r and m, $v^n_0 < v^n_1 < v^b_{r,m}$.*

Our remaining goal in this subsection is to describe key policy rules for banks throughout the state space. We divide our analysis into three propositions regarding meetings with other banks and with nonbanks. First, we describe the behavior of *liquid* banks. Such banks have reserves that exceed their total amount of money in circulation. Since future payoffs are discounted, and $R = 0$, a liquid bank will always find it optimal to issue an additional unit of money to a nonbank in order to consume now.[10] In addition, if a bank's reserves

[10] That behavior does not necessarily follow in meetings with other banks. In the next section, we show simulations in which there is a "multiplier effect." If notes issued to nonbanks have a high float, then some banks may direct new issue towards nonbanks only, avoiding the short-term cost resulting from immediate

sufficiently exceed the amount of its money in circulation, a bank will find it optimal to stop building additional reserves. This is because the benefit from extra reserves will come in the distant future and is, thus, discounted, while the cost of production is borne today.

Proposition 4. *For all r and m, with $r > m$, $\phi^n_{r,m} = 1$. Moreover, for every m, there exists an r_m such that, if $r \geq r_m$,* then *$\gamma_{r,m} = 0$.*

Next, we describe the behavior of *illiquid* banks. These are banks whose total amount of money in circulation exceeds their reserves. When r is relatively low, such banks are facing a positive probability of closure. Note that, as r grows, a bank issuing money at state (r, r) faces a diminishing probability of being caught with negative reserves. In the limit, this probability becomes arbitrarily small. In this case, the short-term gains dominate and the bank issues money, even to other banks, in order to consume. Of course, some illiquid banks are caught with negative reserves and have to exit the sector. To avoid exit, such banks typically find it optimal to suffer the disutility of production in order to improve their reserves.

Proposition 5. *There exists $\hat{r} > 0$ such that, if $r \geq \hat{r}$* then *$\phi^n_{r,r} = 1$.*

The next proposition describes the behavior of banks that are "too illiquid" in the sense that $m - r$ is positive and large. Banks that have very low reserves compared to the amount of their money in circulation will concentrate on issuing more notes instead of building additional reserves, since they expect to exit the banking sector with high probability in the near future. We call such banks *wildcats*.[11]

Proposition 6. *For every r there exists m_r such that, if $m \geq m_r$ then $\phi^n_{r,m} = 1$ and $\gamma_{r,m} = 0$.*

---

redemption.

[11] The term "wildcat bank" originates in the 1800s, when it was applied to banks that were established in inaccessible locations in order to avoid redemptions of their notes. The term has become synonymous with unsound banking practices regarding note issue and the obligations of banks regarding redemptions of their notes.

{Insert Figure A}

The statements in the above propositions are summarized in Figure A (see appendix). As the values of r and m vary, the optimal strategies give rise to the possibility of four regions in the (r, m)-space. In region I reserves are high compared to money in circulation. In that case, a bank finds it optimal to issue money. At the same time, such a bank rejects opportunities to increase reserves. Banks in region II still find it optimal to issue money, but being less liquid, they also accept trades that increase reserves. Banks in region III are becoming alarmingly illiquid and find it optimal to both improve their reserves and stop issuing new notes. In other words, concerned about the possibility of having to exit, these banks concentrate on improving their reserve position.[12] Finally, banks in region IV have too few reserves compared to their notes in circulation and will thus have to exit the banking sector with high probability in the near future. These banks would not benefit from a single-unit increase in their reserves, and thus issue money until they are caught with a negative balance. Notice that the redemption process reduces both reserves and the notes in circulation by the same number and thus always moves a bank southwest, in parallel to the 45-degree line. We remark that banks might enter region IV either accidently, when the redemption process brings them there from another region, or intentionally, after issuing too many notes. In both cases, they never leave this region before exiting the banking sector.[13]

4 Numerical findings

In this section, we present the outcomes of numerical simulations of the model, for several different parametrizations, including some policy experiments. We emphasize that these simulations are not meant to generate quantitative predictions. Rather, they serve as an additional device for exploring the qualitative effects of varying the economic environment on optimal bank behavior. Some of the changes might be interpreted as the result of government

[12] Region III might be a subset of the 45-degree line.

[13] It can be shown that $\phi^n_{r,m} = 0 \Rightarrow \phi^b_{r,m} = 0$, and that $\phi^b_{r,m} = 1 \Rightarrow \phi^n_{r,m} = 1$. In general, even when $R = 0$, the possibility that $\phi^b_{r,m} = 0$ for $r > m$ cannot be ruled out.

policy. We start by describing the equilibrium bank strategies, throughout the state space, of a typical inside-money economy. Then, we study the effects of paying interest on reserves and those of varying the stock of outside money. Finally, we study the effects of different regulations, changing the endowment of initial reserves and prohibiting money creation, followed by parameter changes that increase entry into the banking sector.

Before presenting the simulation results, we shall briefly comment on the assumptions of our model that allowed the numerical task to be reasonably simple. We have assumed that goods are indivisible and money holdings are restricted to $\{0, 1\}$, so that there is no need to compute the *intensive margin*; that is, how much output is traded in each monetary transaction. As indicated in our propositions about the WZ model, if we had ruled out bank failure, and restricted ourselves to outside-money economies, then the linearity of v^b with respect to the interest-rate R would still hold with divisible goods. In that case, it is conceivable to use numerical methods to compute steady states with divisible production and general currency holdings. When we allow for bank failure, however, that linearity is lost, and the state space for banks must include both reserves and notes in circulation. Moreover, predicting then higher moments of the redemption process becomes more complicated, if nonbanks can hold multiple units, and banks have to remember how many notes they issued to each nonbank in the past, now classified by wealth levels. Therefore, assuming the set $\{0, 1\}$ for nonbank holdings greatly simplifies each relevant history to a simple pair, (r, m). Having done that, the assumption of indivisible production further simplifies the analysis. We conclude, therefore, that it might be difficult to extend our experiments with bank failure to general holdings, unless tractable approximations of the much larger state space are developed.

The numerical examples below are based on the following parameter values: we set the number of types, k, to 3; the exogenous probability of death, δ, is fixed at .2; the fraction of newborn agents holding government currency, μ_n, and those born as banks, μ_b, are set at .2 and .1, respectively; utility associated to consumption, u, is fixed at .5 and the disutility from production, e, is set at .2. The above choices for parameter values imply that the sufficient condition in Lemma 1 is satisfied as long as $\beta > .741$. We set the effective discount factor as $\beta = .995$. We also force banks to have a maximum of ten units of reserves and ten units of money in circulation ($N = 10$).

We assume initially that banks are born with two units of reserves ($\bar{r} = 2$), and that no interest is paid on reserves ($R = 0$). Unless stated otherwise, the parameter values remain constant in the examples below.

4.1 The benchmark: An inside-money economy (Example 1)

The parameters and policy settings of *Example 1* define our benchmark. Under this parametrization we find illiquid banks in our simulations. That is, we find banks crossing the 45-degree line in the (r, m)-plane. However, banks still exhibit conservative behavior regarding issue: there are four states on the diagonal $((0,0), (1,1), (2,2), (3,3))$ for which banks do not issue to nonbanks. Moreover, there are states in which a bank, concerned about redemptions, would issue to a nonbank but not to a bank (states $(4,4)$ and $(5,5)$). Note also that there are states where banks exhibit wildcat behavior. In particular, Figures 1 and 2 indicate that there are states in the northwest corner in which illiquid banks do not produce while they choose to issue money. Thus, banks in that region exit the banking sector in the near future. To document wildcat behavior by some banks in equilibrium, it remains to be shown that a positive fraction of banks will enter that region. Figure 2 demonstrates that this is indeed the case. Interestingly, banks do not enter into the wildcat region unwillingly. Rather, banks enter that region because of their decision to print an extra note, and not because of the redemption process. In fact, the redemption process alone will not bring a bank into that region since it moves a bank southwest on a negative 45-degree line (see Figures 1 and 2). Notice that it is possible for a bank to enter the wildcat region even if no note is redeemed during his lifetime up to that point. This is exhibited, for example, by the behavior of a bank that has four units of reserves and keeps issuing notes to nonbanks with no notes being redeemed. Eventually the bank gets to the state $(r, m) = (4, 8)$, where he will enter into the ergodic wildcat region.

{Insert Figures 1 and 2}

4.2 Paying interest on reserves (Example 2)

Illiquidity has non-trivial implications for aggregate welfare in the economy of *Example 1.* Illiquid banks that produce, or even wildcats that do not, play a positive role in providing scarce liquidity. However, there is also a social cost associated to wildcats since these banks stop building reserves (producing) and eventually fail. One possible lesson from *Example 1* is that additional policies might be required for limiting financial fragility, in the simple context of our model. In order to assess the role of bank returns in the matter of stability, we now consider positive interest payments on reserves; that is, $R > 0$. The interest payments are financed with a lump sum tax on banks and nonbanks, deduced from utility in our welfare measures. We consider three values for the interest rate on reserves, $R = 0, .002, .02$ (the economy with $R = 0$ corresponds to *Example* 1).

Table 1 illustrates how a policy of "high" interest rates on reserves ($R = .02$) can be disastrous for nonbank welfare. We find that, as R increases from 0 to .02, average nonbank welfare decreases from 4.54 to 2.911, but average welfare of banks, not surprisingly, increases from 6.9 to 28.09. Because of the interest payed, banks become too concerned about losing reserves and choose not to issue to nonbanks while they are liquid (see Figures 3 and 4 below). Banks thus provide very little liquidity in the economy (they actually destroy liquidity), so the mass of consumers holding notes decreases significantly, relative to *Example 1*, which explains the decline in nonbank welfare.

Table 1 also illustrates general-equilibrium effects that may seem striking. The policy of increasing interest rates, from 0 to .002, raises average nonbank welfare, and decreases average bank welfare (recall that the welfare effects of increasing R from 0 to .02 had the opposite signs). As R is raised to .002, the fraction of the population per type that is willing to spend a note, q, increases from .1083 to .1159 (see Table 1). Nonbank consumers thus have a higher ability to spend in the economy with $R = .002$, which allows them to increase average consumption and welfare (despite the lump sum taxes payed to finance the policy). The reason why nonbanks can buy consumption goods more frequently when $R = .002$ is that private note issue to nonbanks is relatively high in this economy ($p^b \cdot \phi^n$ is .0318 when $R = .002$, instead of .0279 in *Example 1*).

These findings raise the following question: why are banks more willing to issue notes to nonbanks in the economy with $R = .002$ than in the economy with $R = 0$? In principle, when the central bank pays interest on reserves, one would expect banks to be more conservative regarding note issue and to work harder for accumulating reserves. General-equilibrium forces, however, can reverse the sign of these effects. When a low interest rate is used ($R = .002$), banks compete harder for reserves, as illustrated by the observation that there are three states, below the 45-degree line, in which banks do not issue notes to other banks (states $(1,0),(2,0)$ and $(2,1)$ in *Figure 6*). This competition for reserves decreases the consumption frequency of banks, without improving aggregate reserves of the banking sector (when a bank issues a note to another bank, there is just a transfer of one reserve from the bank that issues the note to the bank that receives the note). In particular, one finds that banks are less willing to consume from other banks when $R = .002$, since $p^b \cdot \phi^b$ is now .0064 instead of .0279. This effect translates into a reduction in the welfare of the average bank. A reduction of the average value of banks, in turn, makes them more willing to issue notes as the penalty for exiting the industry becomes smaller. These nonlinear effects, in general equilibrium, lead to an increase of liquidity and to an increase of the average nonbank welfare, which, again, reduces the penalty for exiting the industry and reinforces the incentives to issue more notes. As a result, as R is raised from 0 to .002, the fraction of illiquid banks increases from .003% to 3.6%, with banks producing less (when R is set to .002 banks stop producing in states $(0,3),(1,4)$ and $(2,5)$).

Table 1: Effects of paying interest on reserves			
R (*interest on reserves*)	.0	.002	.02
v_0^n	4.423	4.619	2.824
v_1^n	4.835	5.020	3.267
Avg. nonbank welfare	4.54	4.731	2.911
Avg. bank welfare	6.90	6.814	28.09
p_0^n	.2196	.2167	.2412
p_1^n	.0804	.0840	.0588
$p = p_0^n + p^b \cdot \gamma^b$	.2529	.2492	.2745
$q = p_1^n + p^b \cdot \phi^n$	.1083	.1159	.0588
$p \times q$	.0274	.0289	.0161
$q^b = p_1^n + p^b \cdot \phi^b$	.1083	.0905	.0588
$p^b \cdot \phi^n$	.0279	.0318	.0000
$p^b \cdot \phi^b$	.0279	.0064	.0000
$p^b \cdot \gamma^b$	.0333	.0326	.0333

The following table compares the two environments discussed previously. An interest rate of .02 gives rise to no computed illiquidity, while an interest rate of .002 produces a simulation with a larger mass of illiquid banks than in the economy with $R = 0$.

Table 2: Bank behavior and the interest rate			
	$R = 0$	$R = .002$	$R = .02$
No. states with no consumption	10	7	136
when meeting a bank	6	6	68
when meeting a nonbank	4	1	68
No. states with no production	25	28	15
Fraction of illiquid banks	.003%	3.6%	0%
Fraction of wildcats	.0002%	.16%	0%
No. of banks exiting industry	.0002%	.44%	0%

4.3 Eliminating failure through an injection of reserves (Example 3)

We now return to the case $R = 0$ and investigate the effects of endowing banks with more reserves. As mentioned earlier, injecting more initial reserves is equivalent to allowing a weaker shut-down rule. It can also be loosely interpreted as a lending of last resort for banks with difficulties. We emphasize that in our model, the central bank (or clearing authority) needs to know only the reserve balances of banks and not the number of their notes in circulation.

We compare equilibrium for $\bar{r} = 2$, our benchmark, with $\bar{r} = 4$. The corresponding decision rules are represented graphically in Figures 7 and 8. We find that banks behave more conservatively when they can borrow (are endowed with) 4 units of reserves rather than 2. In fact, banks, in the computed range of $r, m \leq 10$, do not become illiquid when $\bar{r} = 4$ (they only issue notes in states in which $r > m$, both in meetings with banks and with nonbanks). Thus, banks do not risk failure in this simulation. Moreover, banks are also concerned about building reserves. Notice that the number of states in which they decide to produce, in order to improve their reserve position increases.[14] By injecting money into the banking sector, the central bank increases trade and bank values. Consequently, their reserve-management policy becomes more conservative. This result is suggestive of policy recommendations proposed in historical episodes of great liquidity shortages and banking crises.

{Insert Figures 7 and 8}

Some additional findings are summarized in the following table. The table reveals that the liquidity injected by the central bank is absorbed by the banks, so that the mass of nonbanks with money does not increase.

[14] The total liabilities of the central bank are given by all the notes in circulation. Notice that in steady state, some of the notes circulating were issued by banks that have died or have exited the industry. Since the central bank accepts all notes as valid reserves, there is a sense in which the liabilities of the central bank exceed the sum of the liabilities of banks.

Table 3: Effects of an injection of reserves		
	$\bar{r}=2$	$\bar{r}=4$
v_0^n	4.423	4.621
v_1^n	4.835	5.023
Avg. nonbank welfare	4.54	4.733
Avg. bank welfare	6.90	8.147
p_0^n	.2196	.2165
p_1^n	.0844	.0835

As Table 3 shows, both banks' and nonbanks' utility increases when banks receive more initial reserves (are allowed to borrow some reserves). Notice that the increase in utility is larger for banks than for nonbanks. Thus, the penalty for banks exiting the industry (as defined by the difference between the value function of a bank and a that of nonbank consumer) is larger in this case.[15] As a result, banks behave conservatively even in this case of less stringent reserve requirements. The experiment suggests that there is excessive bank failure in *Example 1*, and that the injection of reserves produces an (virtually) outside-money economy with an increase in welfare. The improvement supported by outside money cannot be interpreted, however, as a counter-example to the findings of CW and others. The comparison between inside and outside money is more meaningful when restrictions on money holdings are symmetric across banks and nonbanks. Here, nonbanks can hold at most one unit of money. When banks fail and become nonbanks, there is a welfare loss resulting from the fact that the failed bank cannot store high levels of wealth as before. That gives an advantage to outside money in our simulation. *Example 3* just shows that there exist inside-money economies with excessive failure. In order to add further perspective into the gains provided by inside money, we consider now a liquidity shortage, and how allowing bank illiquidity can improve welfare significantly.

[15] We also find simulations, not included, in which setting β to a high value produces a similar effect: the value of a bank increases relative to that of a nonbank, virtually eliminating bank illiquidity.

4.4 Inside money as a source of liquidity (Example 4)

In this experiment, we modify our benchmark so that outside money is more scarce. To this end, we decrease the fraction of newborn individuals to $\delta = .1$ (instead of $\delta = .2$), and we decrease the fraction of newborn nonbanks with money to $\mu_n = .1$ (instead of $\mu_n = .2$).[16] As a result, the measure of newborn nonbanks with money is reduced by one fourth relative to the benchmark (the product $\delta\mu_n$ is .01 instead of .04). We consider four cases of banking environments, keeping $R = 0$. We vary $\bar{r}$ in $\{1, 4\}$, and consider two possible regulations: one imposing a 100% reserve requirement (imposing outside money), and the other allowing banks to become illiquid (to provide inside money), as before.

We look first at the regulation allowing inside money. Since liquidity is potentially scarce with $\delta = .1$ and $\mu_n = .1$, banks have a hard time building reserves to support issue and, as a result, they are quite willing to become illiquid. The decisions for banks regarding issue are represented graphically in Figure 9, with $\bar{r} = 1$ and $\bar{r} = 4$ represented on the left and right panels, respectively. Figure 9 reveals that issue is indeed less conservative now than in the benchmark. There are now fewer states in the diagonal for which banks do not issue notes. Surprisingly, there are states in which banks, despite being liquid, forgo a consumption opportunity when meeting another bank . In these states (below the 45-degree line), banks avoid issuing money that will be cleared immediately. They prefer to use the extra reserves to support multiple money issue in the future (a kind of multiplier effect).

{Insert Figure 9}

In order to study the role of inside money as a source of liquidity, we now evaluate how welfare would be affected if banks were not allowed to become illiquid. This outside-money regulation could be implemented by an arbitrarily large (exogenous) penalty to banks caught with negative reserves, or by a central-bank policy that prevents banks from printing their own currency, allowing them to put in circulation only government currency (which is earned as they accumulate reserves). A comparison of columns 1 and 2 in Table 4 reveals

[16] A lower value for δ implies that banks have a higher chance of becoming illiquid since banks'expected lifetime is inversely related to δ.

that allowing banks to become illiquid when $\bar{r} = 1$, can provide a significant boost to the average welfare of both banks and nonbanks. Inside money responds to liquidity scarcity, as indicated by the fact that 18.6% of banks in this economy are illiquid. Moreover, private money, as a fraction of the aggregate money stock, increases from 48.7% to 64% as banks are allowed to become illiquid. The increase in liquidity comes at a cost since 1.7% of banks are shut down each period, due to their negative reserve position.[17]

When evaluating the role of banks in providing liquidity, it is helpful to distinguish between inside money and private money, although that may not be a trivial task, as our literature review suggests. In our model, inside money certainly takes the form of note issue not backed by reserves, so that banks create liquidity when they issue unbacked liabilities. Here, when banks issue notes that are backed by accumulated reserves, they are just changing the physical aspects of the means of payments, without necessarily adding additional liquidity to the economy. How much extra liquidity is generated by inside money is an open question, and the choice of an appropriate measurement is not trivial.[18]

Table 4 reveals that, when $\bar{r} = 4$, average welfare for banks and nonbanks also improve when banks are allowed to print their own money (compare columns 3 and 4). We also notice that inside money increases welfare by a larger amount in the economy with low $\bar{r}$. Inside money plays a less prominent role in the economy with $\bar{r} = 4$ because liquidity is less scarce than when $\bar{r} = 1$. When banks are born with four units of reserves, they have an easier time printing money. This, in turn, has positive external effects on other banks that can now accumulate reserves more easily. Because the value of being a bank, relative to a nonbank, is quite large in this economy, banks behave more conservatively, and the fraction of illiquid banks is only 3.5% (instead of 18.6% in the case where $\bar{r} = 1$). Our findings illustrate that

[17]The fraction of failed banks that become wildcats intentionally (issue one more unit), as opposed to accidentally (due to an unlucky outcome of the redemption process), is .9993 for $\bar{r} = 1$, and .995 for $\bar{r} = 4$.

[18]In steady state, aggregate notes issued by banks to nonbanks is measured by $\Delta = k\ p_0^n \sum_{r,m} p_{r,m}^b \phi_{r,m}^n$. Since money is destroyed at a rate π, the stock of circulating notes issued by banks is given by Δ/π. The steady-state creation of inside money can be defined as $\Omega = k\ p_0^n \sum_{r \leq m} p_{r,m}^b \phi_{r,m}^n$, and the stock of inside money measured as Ω/π. In *Example 4*, when banks are allowed to become illiquid in the economy with $\bar{r} = 1$, the measure of inside money, relative to total holdings, increases from 0% to 28.2%.

inside money plays an important role when liquidity is scarce; an observation that we believe is consistent with some historical episodes.[19]

Table 4: Inside money under scarce liquidity ($\delta = .1$, $\mu_n = .1$)

	$\bar{r} = 1$		$\bar{r} = 4$	
Reserve Regulation	*Outside M.*	*InsideM.*	*Outside M.*	*Inside M.*
Avg. nonbank welfare	3.13	4.31	4.35	4.51
Avg. bank welfare	3.95	6.10	7.08	7.61
p_1^n/p_0^n	.200	.319	.314	.335
Number of banks	.10	0.085	.10	.0978
Fraction illiquid banks	0%	18.6%	0%	3.5%
Fraction exiting banks	0%	1.71%	0%	.22%
Share of private money	48.7%	64.0%	64.2%	65.8%
Share of inside money	0%	28.2%	0%	5.5%

4.5 Increasing entry into the banking sector (Example 5)

One question is to what extent redemption disciplines the issue of money by the banking sector. We address this question by increasing competition for reserves through an increase in banking entry. To this end, we fix $\mu_n = .1$ and compare three economies that differ in the fraction of the newborn agents that are banks ($\mu_b = .1, .5, .90$). Table 5 shows that as the fraction of banks in the population increases from .1 to .90, the probability that a unit of money in circulation is redeemed increases from .126 to .367. It should be noted that the speed at which notes are refluxed is determined by aggregate forces (and not just the number of banks). In our economy, the central bank together with the rule that shuts down banks with negative reserves, is quite effective in disciplining banks. Because banks strive to accumulate reserves for supporting issue, the redemption probability increases with the number of banks. This increase, in turn, reinforces their conservative behavior. Figures 10 and 11 compare banks' decision rules when $\mu_b = .1$ and $\mu_b = .5$. When $\mu_b = .5$, banks prefer not to issue money whenever the prospects of excessive redemptions put them at risk

[19] See Hanson (1979) and Cuadras-Morato and Rosés (1998).

of being closed down. In this economy, there is no illiquid banking. The reflux of notes is thus a strong force disciplining banks.

Table 5 demonstrates that welfare for both banks and nonbanks (across the three examples considered) improves with the number of banks. A larger fraction of banks leads to an increase in liquidity, which makes both banks and nonbanks better off. Welfare is the largest in the economy with the smallest aggregate stock of money, which suggests once again that choosing an appropriate notion of liquidity, in an economy with inside money, is a non-trivial matter. In our model, the welfare of the society depends on both the measure of people willing to produce, p, as well as that willing to consume, q, in a multiplicative way: $p \times q$ (the product is a measure of the frequency of successful trades). In equilibrium, $p = p_0^n + p^b \cdot \gamma$ represents how easily a consumer can spend a note, and $q = p_1^n + p^b \cdot \phi^n$ represents how easily a producer can earn a note. The dependence on the product $p \times q$ demonstrates that liquidity is not sufficiently summarized by the money stock. The amount of money that is created and destroyed is also relevant.

While the number of potential banks is our model is exogenous, one could model free entry in the following way. Suppose, in the spirit of HHW, that in each period there is an afternoon market where a general good can be traded. Moreover, suppose that there is a fixed cost of setting up a communication network with the central bank (e.g., creating a bank involves a resource cost in general goods). Then, we could consider an economy where there is free entry into the banking industry. Individuals would be willing to pay (in units of the general good) an amount that would make them indifferent between being a banker or not. We conjecture that, under certain parameter configurations, a stable monetary system will arise in which banks do not make ex-ante profits. Once banks pay the lump-sum fee, they will face a reserve-management problem that, as before, provides incentives to accumulate reserves and to control note issue.

Table 5: Chartering more banks

Fraction of newborn banks	.10	.50	.90
Redemption probability	.126	.25	.367
Avg. nonbank welfare	4.31	5.86	7.50
Avg bank welfare	6.10	8.90	11.92
p_1^n/p_0^n	.319	.468	1.0
Number of banks	0.085	.50	.90
Fraction of illiquid banks	18.6%	0%	0%
Fraction of exiting banks	1.71%	0%	0%
Share of private money	64.0%	75%	
Share of inside money	28.2%	0%	
Aggregate money stock	.221	.159	.0499

5 Conclusion

We have presented a series of results relating liquidity to monetary exchange. We believe that outside-money models in general, and Walrasian models with simple monetary aggregates in particular, are not likely to generate the kind of diverse phenomena observed in monetary history. Models of money and banking are needed to understand monetary history, and to help designing better central-bank policies. In certain historic episodes, private banks have been fairly described as inherently unstable. One challenge is to build models in which banking adversities receive serious consideration while, at the same time, banks identify liquidity needs and promote decentralized trade.

It should not be surprising that models capable of generating such diversity do so at the cost of some complexity. We recognize that our model is not particularly simple, as is evident from the need to perform numerical experiments in this paper. But we also believe that our model is built on sound foundations and that it addresses some of the concerns expressed in the Kareken and Wallace (1980) volume regarding the lack of models of media of exchange and privately-provided liquidity. Banks in our model are simply agents that can be monitored to some extent, and desirable institutions should emerge to make use of their capacity to create inside money and wealth. This is a basic but fruitful view of the role of banks. However, modeling the coexistence of credit and money has always been difficult.

Early monetary theorists emphasized that competition in money provision could create a stabilizing effect through the reduction in the float of private money. They termed this effect "the law of reflux." Our model fits this emphasis well. We have also discussed how a simple scheme of interest payments on reserves may induce a concern by banks regarding reserve management. We have shown that a monetary economy with float of private notes can, under some assumptions, be duplicated by an outside-money economy without float. When banks can create money, however, there is a welfare gain that bank regulation can exploit. If banks are perfectly monitored, an optimal regulation may remove their concern about reserves. The literature review motivated our assumption of a fixed clearing system, according to which banks are allowed to create liquidity but they also risk failure in case of excessive redemptions. Thus, inside money under imperfect monitoring may lead to some bank failure.

In our model, banks can be profitable without earning interests on reserves, since the capability to create liquidity has private benefits. We studied whether a stable monetary system emerges, as well as whether there is underissue or overissue of notes, and whether some infusion of reserves or an interest payment on reserves by the central bank can reduce monetary instability. The answers proved to be non-trivial, and the model displayed important non-linearities and general-equilibrium effects. We showed that banks may underissue if trading with other banks results in excessive redemptions. Small increases on interest rates payed on reserves may end up reducing banking returns, and increasing risk-taking. Larger interest rates may produce an outside-money economy, with worsened nonbank welfare. In other cases, banks willingly cross a threshold in their reserve-management policy and start overissuing until failure. In summary, monetary stability depends on central-bank polices and liquidity conditions in general. Allowing banks to run some negative reserves, which resembles a central-bank policy of reserve lending, may facilitate reserve management, increase returns, and promote trade and financial stability with a lower rate of bank failure.

6 References

Cavalcanti, R. (2004): A Monetary Mechanism for Sharing Capital: Diamond and Dybvig meet Kiyotaki and Wright. *Economic Theory* **24**, 769-788.

Cavalcanti, R., A. Erosa, and T. Temzelides (1999): Private Money and Reserve Management in a Random-Matching Model. *Journal of Political Economy* **107**(5), 929-945.

Cavalcanti, R., and H. Forno (2004a): Money with Bank Networks. Ensaios Econômicos EPGE n. 545, Getulio Vargas Foundation.

Cavalcanti, R., and H. Forno (2004b): Cyclical Interest Payments on Inside Money. Ensaios Econômicos EPGE n. 546, Getulio Vargas Foundation.

Cavalcanti, R., and N. Wallace (1999a): A Model of Private Banknote Issue. *Review of Economic Dynamics* **2**, 104-136.

Cavalcanti, R., and N. Wallace (1999b): Inside and Outside Money as Alternative Media of Exchange. *Journal of Money Credit and Banking* **31** (August, part 2), 443-57.

Chang, R. (1999): *Comment on* Private Money. *Journal of Money Credit and Banking* **31** (August, part 2), 492-98.

Fisher, S. (1986): Friedman Versus Hayek on Private Money. Review Essay. *Journal of Monetary Economics* **17**, 433-439.

Friedman, M., and A. Schwartz (1963): *A Monetary History of the United States 1867-1960.* Princeton, N.J.: Princeton University Press.

Friedman, M. (1959): *A Program for Monetary Stability.* New York: Fordham University Press.

Gorton, G. (1996): Reputation Formation in Early Bank Notes. *Journal of Political Economy* **104**, 346-397.

Hanson, J. R. (1979): Money in the Colonial American Economy: An Extension. *Economic Inquiry* **XVII**, 281-286.

Hayek, F. A. (1990): Denationalization of Money- The Argument Refined. Third Edition. Institute for Economic Affairs, London.

He, P., L. Huang, and R. Wright (2003): Money and Banking in Search Equilibrium. Mimeo.

(available at the homepage of the FRB conference "Models of Monetary Economies II").

Kareken, J., and N. Wallace (1980): *Models of Monetary Economies II.* Minneapolis: Federal Reserve Bank of Minneapolis.

King, R.G. (1983): On the Economics of Private Money. *Journal of Monetary Economics* **12**, 127-158.

Kiyotaki, N., and R. Wright (1989): On Money as a Medium of Exchange. *Journal of Political Economy* **97**(4)**,** 927-954.

Mills Jr., D. (2001): A Model in which Outside and Inside Money are Essential. Purdue University, Mimeo.

Rolnick, A. J., and Weber, W.E. (1983): "New Evidence on the Free Banking Era." *American Economic Review* **73**,1080—1091.

Rolnick, A.J., and Weber, W.E. (1984): The Causes of Free Bank Failures: A Detailed Examination. *Journal of Monetary Economics* **14**, 267–91.

Schwartz, A. J. (1987): Banking School, Currency School, Free Banking School. In *The New Palgrave: A Dictionary of Money and Finance,* edited by P. Newman, M. Milgate, and J. Eatwell. 1st ed. London: MacMillan Press Limited.

Shi, S. (1995). "Money and Prices: A Model of Search and Bargaining." *Journal of Economic Theory* 67, 593-95.

Tobin, J. (1980): *Discussion of* The Ovelapping Generations Model of Fiat Money. In *Models of Monetary Economies II*, edited by J. Kareken and N. Wallace. Minneapolis: Federal Reserve Bank of Minneapolis.

Trejos, A., and R. Wright (1995). "Search, Bargaining, Money and Prices." *Journal of Political Economy* 103, 118-141.

Wallace, N., and T. Zhu (2003): Float on a Note. Mimeo. (available at the authors' homepages).

White, L. H. (1984) *Free Banking in Britain: Theory, Experience, and Debate, 1800–1845.* Cambridge: Cambridge University Press.

Wright, R. (1999): *Comment on* Inside and Outside Money as Alternative Media of Exchange. *Journal of Money Credit and Banking* **31** (August, part 2), 461-68.

Williamson, S. (1999): Private Money. *Journal of Money Credit and Banking* **31** (August, part 2), 469-91.

Appendix

Here we formally describe the mapping T_{γ,ϕ^b,ϕ^n} governing the law of motion of (p^n, p^b). For states such that $\sum_{r,m} p^b_{r,m} + p^n_0 + p^n_1 = \frac{1}{k}$, the mapping is given by the right side of the equations below, together with the expressions from the equilibrium definition relating p^b, p^n, and π to (γ, ϕ^n, ϕ^b). For nonbanks, we have

$$p^n_1 = (1-\delta)(p^n_1 + q\ p^n_0 - p\ p^n_1) + \frac{1}{k}(z + \delta\mu_c)$$

and

$$p^n_0 = (1-\delta)(p^n_0 - q\ p^n_0 + p\ p^n_1) + \frac{\delta}{k}(1 - \mu_c - \mu_b),$$

where $p = p^n_0 + p^b \cdot \gamma$ and $q = p^n_1 + p^b \cdot \phi^n$. For banks, we have

$$\begin{aligned} z \quad = \quad & k\{\sum_{m\geq 1}\sum_{r=0}^{m}\sum_{i=r+1}^{m} \tbinom{m}{i}\pi^i(1-\pi)^{m-i}x_{r,m} + \sum_{m\geq 0} x_{m,m}\ \phi^b_{m,m}\tbinom{m}{m}\pi^m(1-\pi)^{m-m} \\ & - \sum_{m\geq 1}\gamma_{m-1,m}\tbinom{m}{m-1}\pi^{m-1}(1-\pi)^1 x_{m-1,m}\} \end{aligned}$$

and

$$\begin{aligned} p^b_{r',m'} \quad = \quad & (1-\delta)\sum_{m\geq m'}\tbinom{m}{m-m'}\pi^{m-m'}(1-\pi)^{m'} \\ & \times\{q^b\left[\gamma_{r'+m-m'-1,m}\ \ p^b_{r'+m-m'-1,m} + (1-\gamma_{r'+m-m'-1,m})\ \ p^b_{r'+m-m',m}\right] \\ & +p^n_0\ \ \left[\phi^n_{r'+m-m',m-1}p^b_{r'+m-m',m-1} + (1-\phi^n_{r'+m-m',m})\ p^b_{r'+m-m',m}\right] \\ & +\gamma\cdot p^b\ \ \left[\phi^b_{r'+m-m'+1,m}\ p^b_{r'+m-m'+1,m} + (1-\phi^b_{r'+m-m',m})\ p^b_{r'+m-m',m}\right] \\ & +(\frac{1}{k} - q^b - p^n_0 - \gamma\cdot p^b)\ p^b_{r'+m-m',m}\} + I_{r',m'}\frac{\delta\mu_b}{k} \end{aligned}$$

for all $r', m' \geq 0$, where $q^b = p^n_1 + p^b \cdot \phi^b$, and $I_{r',m'}$ is an indicator function that assumes value one if $(r', m') = (\bar{r}, 0)$ and value zero otherwise.

Proof of Proposition 1: After imposing the stated linearity on v^b, it follows that

$$w^j_{r,m} = rR + \beta v^b_{r,m} - \beta K f(m-j)$$

where $f(m)$ denotes the mean of the binomial distribution on $\{0, 1, ..., m\}$ with parameter π. Substituting now w into the equation for v gives a Bellman equation whose solutions must satisfy

$$v^b_{r+i,m+i} = v^b_{r,i} + a_1 i - a_2[f(m+i) - f(m)]$$

where a_1 and a_2 are constants. Since the mean $f(m)$ of a binomial distribution is linear in m, the stated linearity of v^b follows. The uniqueness of a solution v^b, given (ϕ, γ), follows from the fact that the Bellman equation for banks satisfies the monotonicity and discounting properties, Blackwell's condition for a contraction. The fact that A is positive follows from the fact that R is positive and from the implied values for a_1 and a_2.

Proof of Proposition 2: Allowing for swaps only adds a constant to bank values in the float economy. Hence we can ignore swaps and write, without loss of generality, the Bellman equations for banks with $m = 0$ in the float economy as

$$v^b_{r,0} = \beta v^b_{r,0} + rR + p^n_0 \phi_{r,0}[u + \beta(v^b_{r,1} - v^b_{r,0})] + p^n_1 \gamma_{r,0}[R - e + \beta(v^b_{r+1,0} - v^b_{r,0})].$$

By Proposition 1, $v^b_{r,1} - v^b_{r-1,0}$ is equal to a constant that does not depend on r. It is clear, by the definition of $\bar{R}$, that $\bar{R} = \beta v^b_{r,1} - \beta v^b_{r-1,0}$ must hold. The result thus follows by substituting $v^b_{r,1} = v^b_{r-1,0} + \bar{R}/\beta$ above and by setting $v^b_g = v^b_{g,0}$ to obtain the Bellman equation of the economy without float.

Proof of Proposition 2: The proof is a straightforward version, in discrete time, of the arguments in Wright (1999), and is thus omitted.

Proof of Lemma 2: We fix $R = 0$ and first demonstrate that v^b is weakly increasing in r. Let $G = \{f : \mathbb{N}_+ \times \mathbb{N}_+ \rightarrow \mathbb{R}\}$. Define the operator $T : G \rightarrow G$ as the right hand side of the functional equation for a bank. We know that T has a unique fixed point, v^b, in the space of bounded continuous functions. If T also maps the space of weakly increasing continuous functions into itself, and since this space is complete, this fixed point will also be a weakly increasing continuous function. We need to show T preserves monotonicity, i.e., for any fixed m, $r_1 \leq r_2$ implies that $Tv^b(r_1, m) \leq Tv^b(r_2, m)$. We first show that for all m and any j, $w^j_{r_1,m} \leq w^j_{r_2,m}$. First, consider the case where $r_1 \geq m$. Then $w^j_{r_h,m} = \sum_{0 \leq i \leq m} \binom{m}{i} \pi^i (1-\pi)^{m-i} v^b_{r_h - i, m+j-i}$, where $h = 1, 2$. Notice that the terms multiplying v^b are the same for $h = 1, 2$. Since v^b is monotone, it follows that $w^j_{r,m}$ is monotone. Next, consider the case where $r_2 < m$. Then

$$w^j_{r_h,m} = \sum_{0 \leq i \leq m} \binom{m}{i} \pi^i (1-\pi)^{m-i} v^b_{r_h - i, m+j-i} + \sum_{r_h \leq i \leq m} \binom{m}{i} \pi^i (1-\pi)^{m-i} v^n_1.$$

Note that the first r_1 terms in the first sum are weakly greater when $r_h = r_2$ than when $r_h = r_1$, given that v is monotonically increasing in r. The result then follows since, for all (r, m), $v^b_{r,m} > v^n_1$. By the same reasoning, the desired inequality follows for the case where $r_1 < m < r_2$, since v^b is monotonically increasing in r and, for all (r, m), $v^b_{r,m} > v^n_1$.
Now we show that v^b is weakly decreasing in m. Applying the same reasoning we need to show that for any fixed r, $m_1 \leq m_2$ implies that $Tv^b(r, m_1) \geq Tv^b(r, m_2)$. Like before, we first show that for all r and any j, $w^j_{r,m_1} \geq w^j_{r,m_2}$. We consider the case where $m_2 < r$ first. In that case, $w^j_{r,m_h} = \sum_{0 \leq i \leq m_h} \binom{m_h}{i} \pi^i (1-\pi)^{m_h - i} v^b_{r-i, m_h+j-i}$. Define $p(i, m)$ by $p(i, m) = \binom{m}{i} \pi^i (1-\pi)^{m-i}$. We then have $\frac{p(i,m)}{p(i,m+1)} = \frac{\frac{m!}{(m-i)!i!}}{\frac{(m+1)!}{(m+1-i)!i!}} \frac{\pi^i}{\pi^i} \frac{(1-p)^{m-i}}{(1-p)^{m+1-i}} = \frac{m+1-i}{m+1} \frac{1}{1-p}$. This expression is less than one if and only if $i > p(m+1)$. Therefore, $p(i, m)$ is greater than $p(i, m+1)$ for low values of i and is lower than $p(i, m+1)$ for high values of i. The result then follows since $w^j_{r,m}$ is a convex combination of decreasing functions of m. Therefore, $w^j_{r,m_1} \geq w^j_{r,m_2}$. Since $v^b_{r,m} > v^n_1$, the same argument can be used to demonstrate the result both when $r < m_1$ and when $m_1 < r < m_2$. ■

Proof of Lemma 3: The first inequality clearly holds. Regarding the second inequality, since v^b is increasing in r and decreasing in m, it is sufficient to show that $v^b_{0,m} > v^n_1$, for m large. In this case, the bank will have a negative balance with probability 1 at the end of the period and will thus exit the sector. If it issues one more unit of money this period, in the next period it will have the value of a nonbank with one unit of money, v^n_1. So $v^b_{0,m} > v^n_1$, even for arbitrarily high m. ■

Proof of Proposition 4:

(First part). The proof follows by contradiction. Assume that there is an optimal policy (ϕ^n, ϕ^b, γ) such that $\phi^n_{r_0,m_0} = 0$ for some state (r_0, m_0) with $r_0 > m_0$. Consider a bank that is in state (r_0, m_0). Notice that this bank decides not to issue a note in the current period. Also, notice that this decision would only be optimal if the bank could reach with positive probability in the future a state where it will issue a note (otherwise, the bank would trivially increase utility by consuming today instead of not consuming forever). Denote this state by (r_1, m_1).[20] We shall show that there is a policy $(\hat{\phi}^n, \hat{\phi}^b, \hat{\gamma})$ that gives a strictly higher expected discounted utility to the bank than policy (ϕ^n, ϕ^b, γ), contradicting the optimality of (ϕ^n, ϕ^b, γ). In order to define the new policy $(\hat{\phi}^n, \hat{\phi}^b, \hat{\gamma})$, it is convenient to expand the state space to include two new artificial state variables, representing artificial reserves r_a and notes m_a. Initialize the current (expanded) state as $(r, m, r_a, m_a) = (r_0, m_0, r_0, m_0)$. Set $\hat{\phi}_{r_0,m_0,r_0,m_0} = 1$ and, immediately after the note is issued, let the expanded state variable become $(r, m, r_a, m_a) = (r_0, m_0+1, r_0, m_0)$. From now on, update r_a and m_a in the same way as r and m but with the following exception: If the note just issued in state (r_0, m_0, r_0, m_0) is redeemed, do not count this redemption in the state for artificial reserves and notes (that is, do not decrease r_a and m_a by 1). Set $\hat{\phi}^n_{r,m,r_a,m_a} = \phi^n_{r_a,m_a} \hat{\phi}^b_{r,m,r_a,m_a} = \phi^b_{r_a,m_a} \hat{\gamma}_{r,m,r_a,m_a} = \gamma_{r_a,m_a}$ for all states (r, m, r_a, m_a) but with the following exception. Recall that there exists a state (r_1, m_1) that can be reached with positive probability and where $\phi^n_{r_1,m_1} = 1$. When such a state is reached for the first time, set $\hat{\phi}^n_{r_1,m_1,r_a,m_a} = 0$ and change the law of motion for the artificial state to allow the note that would have just been issued under the original policy in state (r_1, m_1) (but that in fact was issued some periods ago under the alternative policy) to be redeemed with probability π (this is an artificial redemption since the original note could be redeemed earlier). From then on, decisions of the bank are as indicated by the policy $(\hat{\phi}^n, \hat{\phi}^b, \hat{\gamma})$. Now we contradict the optimality of (ϕ^n, ϕ^b, γ) by noting the following:

1. Under the alternative policy, the bank consumes today (when the bank is in state (r_0, m_0)), rather than at some point in the future when the state (r_1, m_1) is reached.

[20] Obviously, since no new notes were issued before reaching state (r_1, m_1), it should be the case that $r_1 > m_1$ for all possible histories of redemptions and production decisions.

2. Consumption and production decisions are the same while the bank transits from (r_0, m_0) to (r_1, m_1). Because, $r_0 > m_0$ and $r_1 > m_1$ (see last footnote), the bank does not risk failure while transiting from (r_0, m_0) to (r_1, m_1) under policy $(\hat{\phi}^n, \hat{\phi}^b, \hat{\gamma})$.

3. Once state (r_1, m_1) is reached, the stochastic process for (r, m) induced by the redemption process and policy (ϕ^n, ϕ^b, γ) is the same stochastic process for (r_a, m_a) induced by the modified redemption process (considering the artificial redemption for the note issued in state $(r_{0,}m_0)$). As a result, the consumption and production decisions implied by the two policies coincide.

Since banks are impatient, it follows that the policy $(\hat{\phi}^n, \hat{\phi}^b, \hat{\gamma})$ gives higher utility, contradicting that (ϕ^n, ϕ^b, γ) is optimal.

(Second part). Consider a bank facing an opportunity to increase reserves. We have that $\gamma_{r,m} = 0$ if and only if $-e + w^0_{r+1,m} \leq w^0_{r,m}$. Note that $w^0_{r,m}$ is bounded, since it belongs to the interval $(0, \frac{u}{1-\beta})$. It is also an increasing function of r. Therefore, for fixed m, we have $\lim_{r\to\infty} w^0_{r,m} = \lim_{r\to\infty} w^0_{r-1,m} = K_m$. The result then follows for r large. ■

Proof of Proposition 5: Consider the case where $m = r$, where r is large. We have that $\phi^n_{r,r} = 1$ if and only if $u + w^1_{r,r+1} > w^0_{r,r}$. Notice that $w^0_{r,r}$ is bounded, since it belongs to the interval $(0, \frac{u}{1-\beta})$. It is also an increasing function of r. Therefore, $\lim_{r\to\infty} w^0_{r,r} = \lim_{r\to\infty} w^0_{r,r+1} = K$, for some finite constant K. Also, $w^0_{r,r+1} < w^1_{r,r+1} < w^0_{r,r}$, which implies that $\lim_{r\to\infty} w^0_{r,r} = \lim_{r\to\infty} w^0_{r,r+1} = \lim_{r\to\infty} w^1_{r,r+1} = K$. Thus, the above inequality follows for r large. ■

Proof of Proposition 6: A bank that is given the opportunity to issue a note to a nonbank faces the following problem:

$$\max_{\phi^n_{r,m}} \left\{\phi^n_{r,m}[u + w^1_{r,m+1}] + (1 - \phi^n_{r,m})w^0_{r,m}\right\}.$$

We have that for any fixed r,

$\lim_{m\to\infty} w^j_{r,m} = \lim_{m\to\infty} \beta \left\{\sum{}_{i=0}^{r} p(m,i) v^b_{r-i,m+j-i} + \sum{}_{i=r+1}^{m} p(m,i) v^n_1\right\}$.

The first sum in the bracket involves a finite number of terms, so we have $\lim_{m\to\infty} p(m,i) = \lim_{m\to\infty} \left\{\frac{m!}{(m-i)!i!}\pi^i(1-\pi)^{m-i}\right\} = \lim_{m\to\infty} \pi^i(1-\pi)^{m-i} = 0$. Thus, $\lim_{m\to\infty} w^j_{r,m} = \lim_{m\to\infty} \{\ 0 + v^n_1 \sum{}_{i=r+}^{m}$

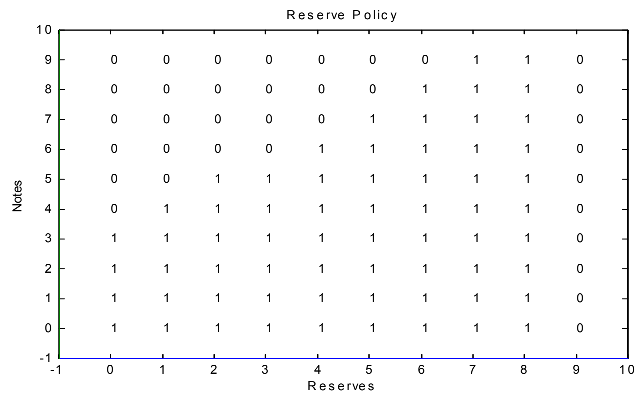

Figure 1: Illiquid Banking and Reserve Policy

v_1^n, and, for any fixed r, there exists an m_r large enough such that $u+\beta v_1^n > \beta v_1^n$ and, therefore, $\phi_{r,m}^n = 1$. We know that $\gamma_{r,m} = 0$ if and only if $-e + w_{r+1,m}^0 \leq w_{r,m}^0$. Now, fix $r \geq 0$. We have that $\lim_{m\to\infty} w_{r+1,m}^j = \lim_{m\to\infty} w_{r,m}^j = v_1^n$. Therefore, given r, there exists a large enough m_r such that $-e + \beta w_{r+1,m}^0 = -e + \beta v_1^n < \beta v_1^n = \beta w_{r,m}^0$, for $m \geq m_r$. Therefore, $\gamma_{r,m} = 0$ for all $m \geq m_r$. ■

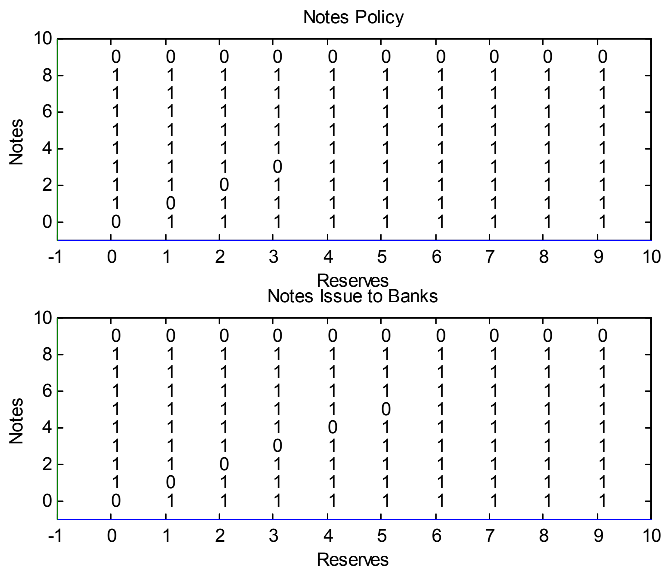

Figure 2: Illiquid Banking and Note Issue

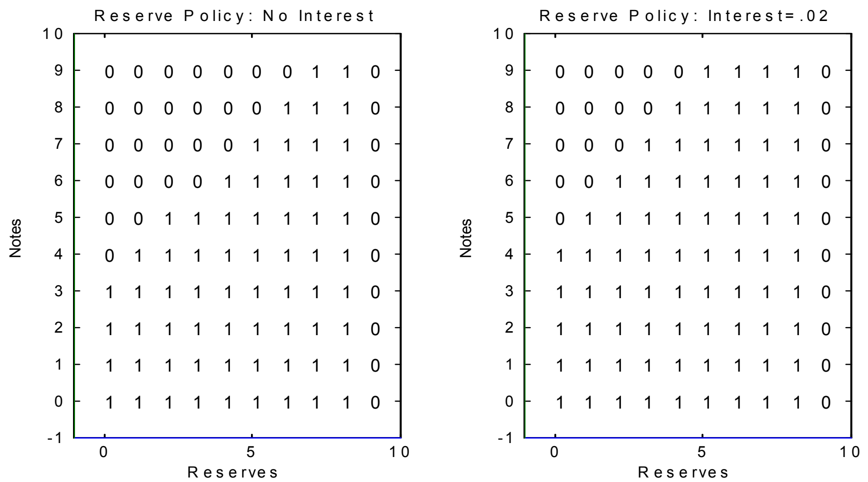

Figure 3: Reserve policy with high interest

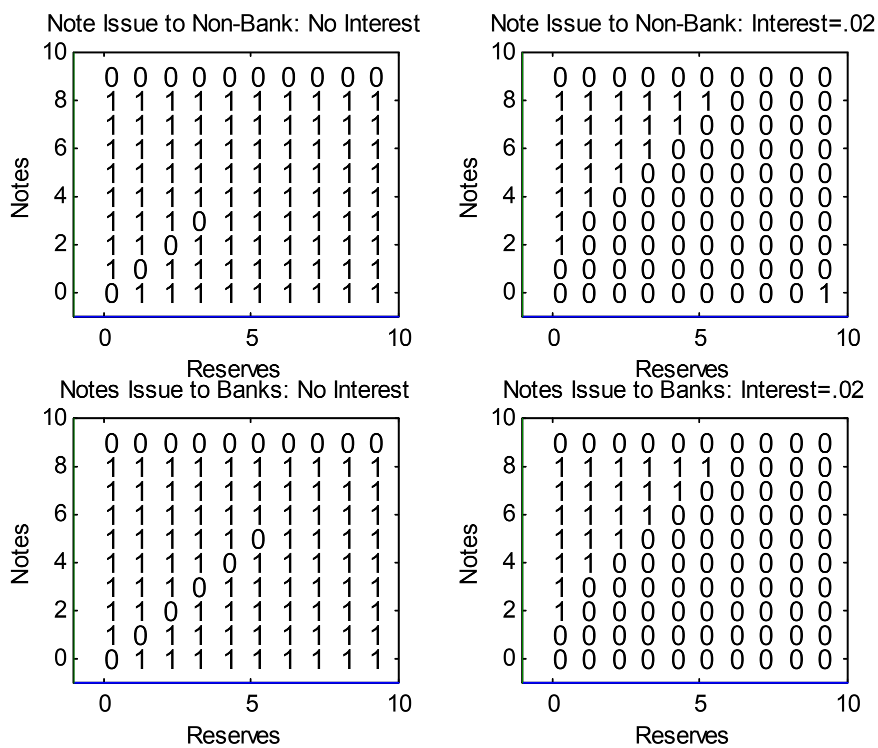

Figure 4: Note issue with high interest

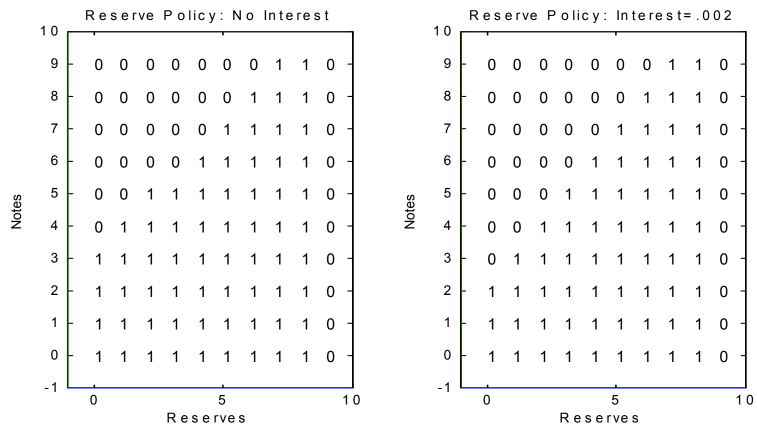

Figure 5: Reserve policy with low interest

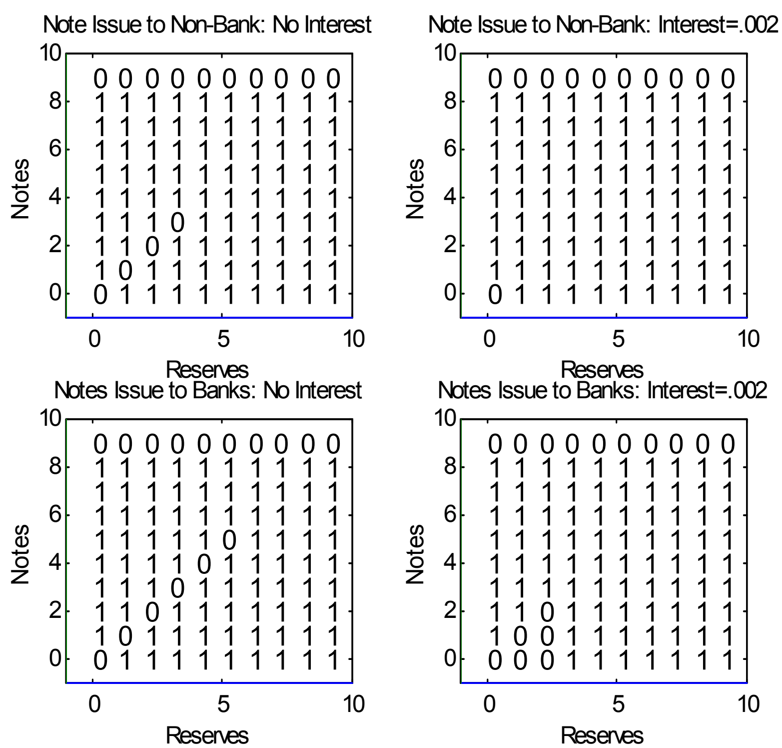

Figure 6: Note issue with low interest

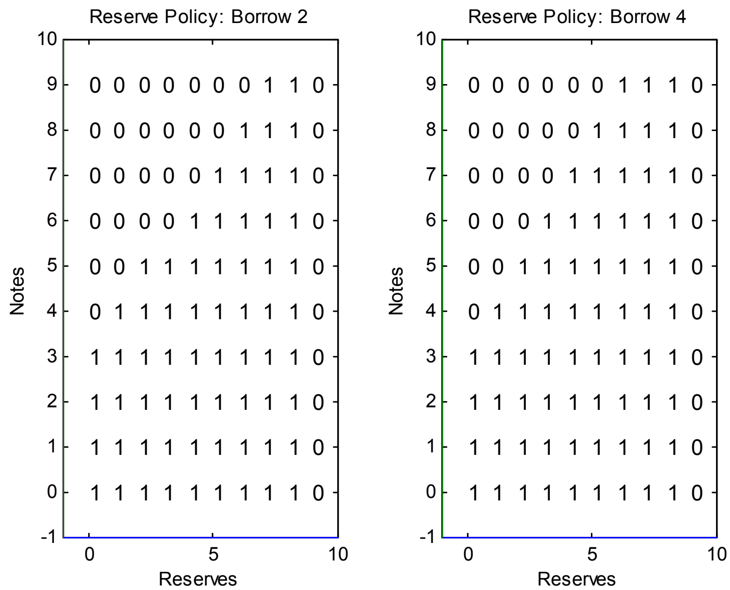

Figure 7: Borrowing and Reserve Policy

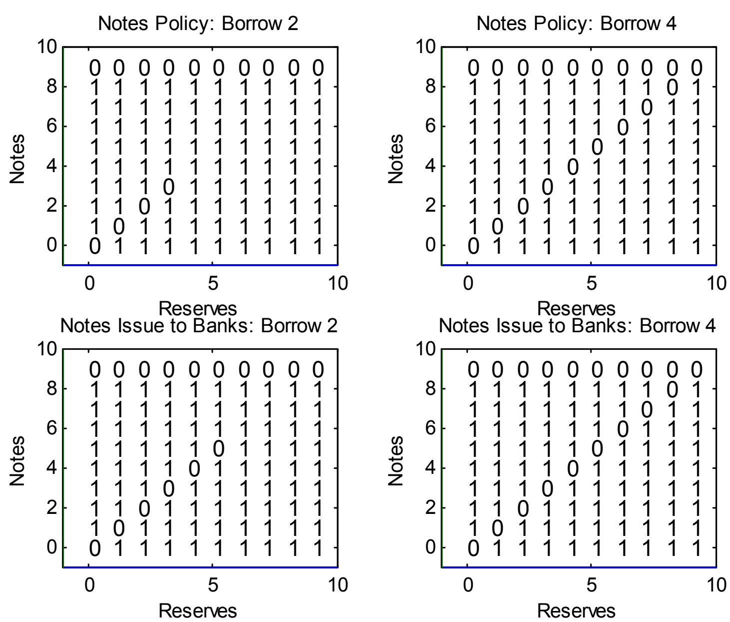

Figure 8: Borrowing and Note Issue

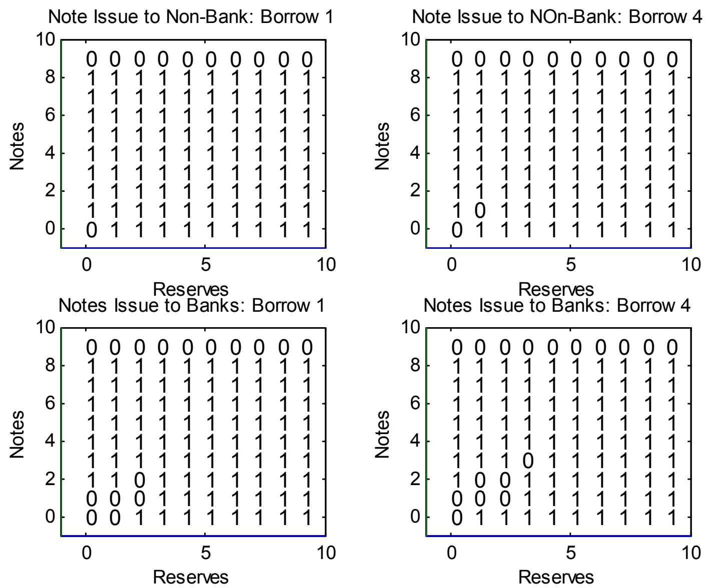

Figure 9: Note Issue with Scarce Liquidity

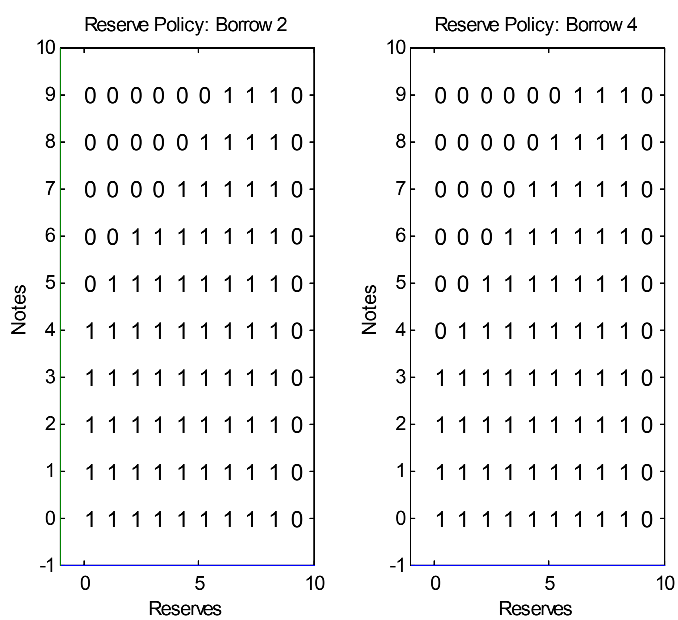

Figure 10: Number of Banks and Reserve Policy

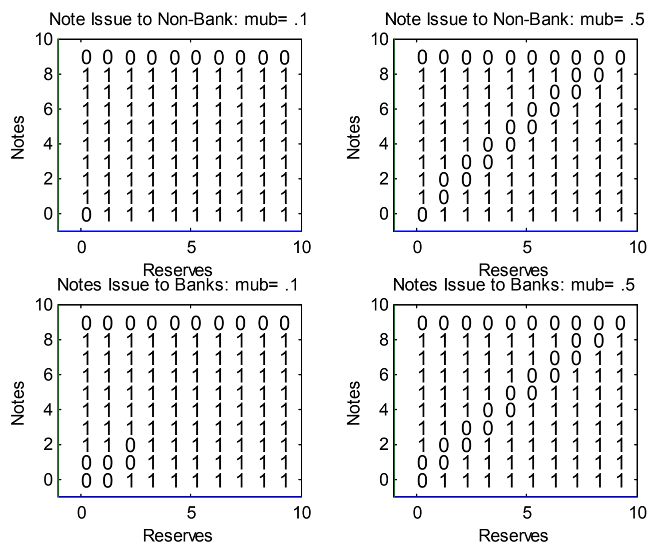

Figure 11: Number of Banks and Note Issue

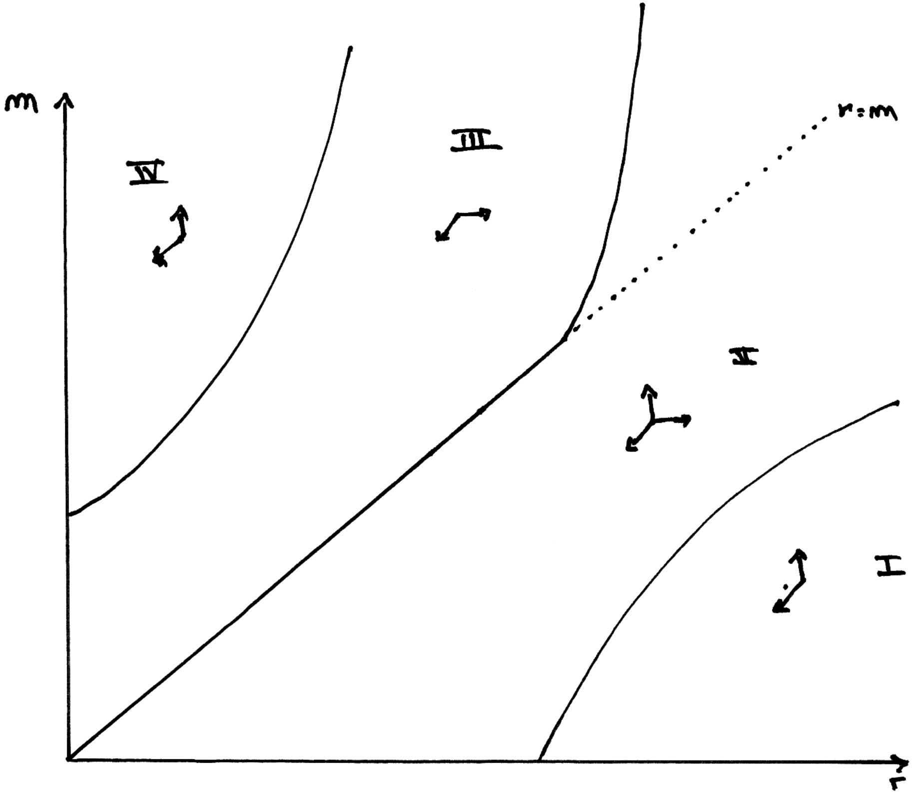

FIGURE A: RESERVE MANAGEMENT