KRANNERT GRADUATE SCHOOL OF MANAGEMENT

Purdue University
West Lafayette, Indiana

Efficient Monetary Allocations and the Illiquidity of Bonds

by

Paola Boel
Gabriele Camera

Paper No. 1171
Date: November 2004

Institute for Research in the Behavioral, Economic, and Management Sciences

Efficient Monetary Allocations and the Illiquidity of Bonds[1]

Paola Boel and Gabriele Camera
Purdue University

November 4, 2004

Abstract

We construct a monetary economy with heterogeneity in discounting and consumption risk. Agents can insure against this risk with both money and nominal government bonds, but all trades must be monetized. We demonstrate that a deflationary policy *à la* Friedman cannot sustain the efficient allocation. The reason is that no-arbitrage imposes a stringent bound on the return money can pay. The efficient allocation can be sustained when bonds have positive yields and—under certain conditions—only if they are illiquid. Illiquidity—meaning bonds cannot be transformed into consumption as efficiently as cash—is necessary to eliminate arbitrage opportunities.

Keywords: Money, Heterogeneity, Friedman Rule, Illiquid Assets
JEL codes: E4, E5

[1] Preliminary version. We thank seminar participants at Purdue, the University of Iowa's Clarence Tow Conference on Money and Banking, in May 2004, the SED 2004 in Firenze, and the monetary conference at the Cleveland Fed in August 2004, for constructive criticism.

1 Introduction

A considerable amount of theoretical work, based on disparate modeling approaches, supports the notion that allocative efficiency in a monetary economy hinges on a deterministic deflationary policy, known as the 'Friedman rule.'[2] The result revolves around the idea that if impatient agents must do business with money *and* must hold cash in their portfolios, then deflation is socially desirable as it lessens the (opportunity) cost of holding cash.

In this theoretical literature, interest-bearing asset trades do not seem to play an essential role in policy execution; lump-sum transfers do the trick. Yet, in practice monetary interventions take the form of open market operations in which cash trades for less liquid interest-bearing government assets. Agents buy bonds when they have excess liquidity, and sell them as the need for transactions balances arises. We attempt to reconcile these observations with the theory, posing the following questions. Does the Friedman rule always lead to a first best? If, not what is the reason and would government liabilities other than money help improve the allocation? If so, should these assets be illiquid? Clearly, the answers hinge on the specification of the economic environment.

We work with a spatially separated model (as in Townsend, 1980) where money has a fundamental allocative role. Agents face random consumption needs but—due to carefully specified environmental frictions (as in Aliprantis, Camera and Puzzello, 2004)—are *physically* prevented from lending or borrowing from each other and from sharing information over time. Thus, trade must be monetized and a sudden consumption shock generates an immediate need for cash. Further, we draw from the two-market formulation of Lagos and Wright (2002) to achieve degeneracy in asset holdings and—as suggested by Berentsen, Camera and Waller (2003)—we work under the assumption of competitive markets.

Two features set apart this model from these related monetary frameworks. First, agents need not rely exclusively on cash to insure against consumption shocks. They can also acquire government nominal bonds, and liquidate them for cash before maturity if a consumption need arises. Second, the model accounts for the possibility of a natural form of heterogeneity in that agents differ in their rate of time preference and in their exposure to consumption risk.

We prove two results. First, we demonstrate that deterministic deflations under zero interest rates—that is under the Friedman rule—cannot sustain the stationary efficient allocation. The reason is, under zero interest rates, agents insure against consumption

[2] E.g. the spatial economy of Townsend (1980) or Williamson (2004), money in the utility function or transaction costs models (Chari et al.,1996), the search models of Shi (1997) or Lagos and Wright (2003).

shocks with cash. Unfortunately, a deflationary policy that gives cash the return desired by the most impatient agents leaves the door open to arbitrage opportunities for everyone else. This constrains the equilibrium deflation to the *lowest* discount rate—much as in Becker (1980)—so that impatient agents will tend to under-insure.

Interestingly, if bonds pay no interest (equivalently, if money is the only asset available), then in equilibrium every agent holds positive cash balances. This is unlike Becker—where the most patient agents hold the entire stock of assets (capital)—because money in our model is essential to execute trades. Of course, the more impatient hold less of it, which is detrimental to trading efficiency. This result appears to be quite general and should hold in any environment with heterogeneity in discounting where money has an explicit medium-of-exchange role.

Then, we show that the efficient allocation can be sustained when bonds pay a positive yield but—under certain conditions—*only if* bonds are sufficiently illiquid. That is to say, only if bonds cannot be transformed into consumption as quickly and efficiently as cash. If the government prices bonds correctly, agents fully insure against consumption shocks by holding illiquid bonds that are sold for cash once a consumption need arises. In short, we need a friction (illiquidity) to cure an inefficiency, much as in Kocherlakota (2001).

When is illiquidity a *necessary* friction? When the most patient agents are also those who experience the greatest incidence of consumption shocks. Illiquidity acts as a proportional tax that lowers the bond's expected return according to the anticipated incidence of consumption shocks. Thus, illiquidity affects the bonds' expected return unequally across agent types. By selecting appropriate bond yield and illiquidity parameters, the policymaker can therefore manipulate the rates of return in order to induce agents to perfectly insure against consumption risk while removing possible arbitrage opportunities.

2 Related Literature

Our work complements the literature concerned with the optimality of the Friedman rule. Among such papers is Faig (1988) who finds that the Friedman rule is optimal in a shopping-time model where money reduces transactions costs. Freeman (1993) confirms this result for a model of infinitely lived agents deriving utility from holding real balances. However, in overlapping generation models with dynastic preferences and no intergenerational transfers, the Friedman rule is optimal only if bequests are positive.

Zero nominal rates are also optimal in Williamson's (1996) cash-in-advance, sequential markets model, unless there are money demand (preference) shocks. In that case benefits arise from an accommodating monetary stance so the Friedman rule can be suboptimal. In

Paal and Smith's (2000) growth model shocks to agents' "liquidity preferences" create an insurance role for banks. At high nominal interest rates banks economize on reserves and provide less liquidity, while at low rate they hold higher reserves but fund little investment in capital. The trade-off between liquidity provision against higher growth implies that the Friedman rule is suboptimal when the maximal rate of real growth it can sustain is at or below the real interest rate. Random preferences over future consumption also imply suboptimality of the Friedman rule in Berentsen and Strub (2004).

Phelps (1973) argues against deflationary policies from a pure optimal taxation perspective, noting that the Friedman rule leads to increasing other (more distortionary) taxes, to replace the lost revenue. This contrasts with Kimbrough's (1986) finding that the Friedman rule is optimal even under distortionary taxes since—he argues—money is an intermediate good and as such it should not be taxed.

da Costa and Werning (2003) also suggest the Friedman rule can be optimal under distortionary taxation. So do Chari, Christiano, and Kehoe (1996), in a cash-credit model, a money-in-the-utility-function model and a shopping time model. Their conclusion is confirmed by Correia and Tales (1996)—given distortionary taxes—but is questioned by Mulligan and Sala-i-Martin (1997) who argue that some assumptions in Chari et al (1996) are not based on micro-founded theories of money demand, and give examples that overturn their results. In addition, they conduct an empirical analysis suggesting that in the U.S. the optimal inflation rate is small, but positive.

We, too, consider different nominal interest rules but—despite having non-distortionary taxes—a deflationary policy is not a first best in our model. However, it unambiguously generates beneficial effects. In short, the Friedman rule is always a second best.

Our paper contributes also to research concerned with the allocative role of government-supplied illiquid assets. For instance, Woodford (1990) shows that government borrowing has beneficial effects, for example it smooths endowment fluctuations, when some face illiquidity problems (borrowing constraints). Kocherlakota (2001) argues that government illiquid bonds are essential, when unobservable preferences hinder a more efficient allocation of consumption. Such bonds allow cash transfers from agents with less to more pressing consumption needs. Shi (2003), studies how interest rates and output depend on bonds' endogenous illiquidity (the fraction of unmatured bonds held by buyers). He finds that illiquid bonds can yield higher welfare and higher nominal interest rates.

We too motivate disparities in desired cash holdings via unequal preferences. However, we have no search but competitive markets, unlike Shi; unlike Kocherlakota, private information does not play a role and the Friedman rule is suboptimal. Our notion of illiquidity also differs, as in our model bonds cannot be *directly* exchanged for goods but can be readily cashed in, at a cost.

3 The Model

We consider a discrete-time production economy with a unit continuum of heterogeneous infinitely-lived agents. They are subject to idiosyncratic trading risk and participate in an infinite sequence of markets characterized by spatial separation and enforcement limitations. We adopt these modeling features as they generate an explicit transactions role for currency. However, they also generate complicated history-dependent distributions of assets (e.g. Camera and Corbae, 1999). To ease these complications, we specify a physical environment based on the meeting technology formalized in Aliprantis, Camera and Puzzello (2004), and we draw from the trading and preference formulation of Lagos and Wright (2003).

3.1 Physical Environment

The physical environment is in the tradition of Townsend (1980). It has features that preclude borrowing and lending among agents, while giving an explicit role to money in facilitating spot exchange.

At each date countably many trading groups are formed (think of these as islands). Each group consists of an identically large number of agents, with an identical proportion of agent types. Trading groups are spatially and informationally separated in that an agent can only interact and communicate with his current groupmates. Everyone is in some trading group at any point in time, but obstacles to the flow of information and resources preclude borrowing/lending among agents. Precisely, agents stay in the same trading group only once and move across groups in a way that severs all possible direct and indirect links among traders. Especially, any two agents meet only once and their histories evolve in such a way that the members of any trading group are anonymous. That such a theoretical construct exists, is proved in Aliprantis et al. (2004).

As in Lagos and Wright (2003), two goods markets open and close sequentially on each island, at each date. As in Berentsen, Camera and Waller (2003), the markets are competitive. A different perishable good can be produced on each market, a specialty good in market one, and a general good in the other. Also, the first market is characterized by idiosyncratic trading risk, while the second is not. Specifically, we assume two types of agents $j = H, L$ in proportion ρ and $1 - \rho$. These agents differ as follows. The discount factors β_j satisfy $0 < \beta_L < \beta_H < 1$ and the probabilities α_j that a type j trades in the first market satisfy $0 < \alpha_L < \alpha_H \leq 1$. These trading shocks are drawn at the beginning of each period. Once on market one, consumption and production are equally likely and mutually exclusive, so a trader either desires the good or can produce it; this generates idiosyncratic consumption risk. In short, the more patient agents are more actively involved in trade but are also more actively exposed to consumption risk.

As soon as the first market closes, the second market opens in which everyone participates by producing and consuming. Thus, while only $\rho\alpha_H + (1-\rho)\alpha_L$ of the population trades in market one, everyone trades in the second market, and there is always an equal number of buyers and sellers in each market. Hence, there is no aggregate trading risk. As in Lagos and Wright (2003), it is assumed that those who desire a specialty good derive utility $u(c)$ from $c \geq 0$ consumption, while producers of $c_s \geq 0$ specialty goods suffer disutility $-c_s$. Agents derive utility $U(q) - q_s$ from $q \geq 0$ consumption and $q_s \geq 0$ production of the general good. The functions u and U satisfy the standard Inada conditions and $u(0) = U(0) = 0$. Also, let c^* be the solution to $u'(c) = 1$ and let q^* be the solution to $u'(q) = 1$.

3.2 Assets and the Government

Given the structure of the model, specialty goods trade must be monetized. We assume a government is the sole supplier of fiat currency available in the amount $\mathcal{M}_t > 0$ at the beginning of date t. We let $\mathcal{M}_t = \pi\mathcal{M}_{t-1}$ be the deterministic law of motion of the money stock. As in Lucas (1980), we assume lump-sum cash transfers/taxes, denoted T_t, to keep the announced rate of growth constant. These occur in the second market so $\mathcal{M}_t$ cash is available in market two of period $t-1$.

The government also buys and sells one-period nominal bonds having two distinctive features (similar to U.S. Savings bonds). First, they are non-negotiable claims to currency; bonds cannot be directly exchanged for goods and can be redeemed only by their owner. To formalize it, assume bonds are intangible (hence non-transferable) assets, ownership of which is recorded by the government. Of course, the government can credibly commit to repayment, as it can print currency. Second, bonds are illiquid in that early redemption may come at a cost and cannot involve fractions of the asset. Specifically, bonds are issued in market two at price $p_A \leq 1$ and mature the following period (in market two) paying off one unit of money. Unmatured bonds can be redeemed for $p_\ell \leq 1$ money by traders in market one. Hence, p_ℓ naturally captures the notion of *illiquidity* as the cost of immediate execution of a trade: $1 - p_\ell$ is lost to convert a bond into immediate cash, at the beginning of a period.

4 Stationary Monetary Allocations

We start by indicating the timing of events in any period t for an agent type j. He enters t with portfolio $\Omega_{j,t} = (M_{j,t}, A_{j,t})$ listing non-negative amounts of money and bonds bought with cash in market two of $t-1$. Then, the trading shock is realized, denoted $k = n, s, b$, i.e., n if he cannot trade and b or s if he can buy or sell. If he trades in market one he can choose to liquidate his bonds. As soon as market one closes, the agent enters market two with portfolio $\Omega_{k,j,t} = (M_{k,j,t}, A_{k,j,t})$.

We denote $p_{i,t}$ the price of goods in market $i = 1, 2$. In market one we denote $c_{j,t}$ the consumption of a buyer of type j, and denote c_t the production of any seller. In market two, q_t denotes consumption of any buyer, and $q_{k,t}$ production of those who experienced shock k.[3] Of course, liquidation of bonds is desirable only if cash is needed to consume in market one. Thus, without loss in generality we let

$$A_{b,j,t} \in \{0, A_{j,t}\} \quad \text{and } A_{s,j,t} = A_{n,j,t} = A_{j,t} \tag{1}$$

$$M_{b,j,t} = M_{j,t} + p_\ell(A_{j,t} - A_{b,j,t}) - p_{1,t}c_{j,t}, \quad M_{s,j,t} = M_{j,t} + p_{1,t}c_t, \quad M_{n,j,t} = M_{j,t}. \tag{2}$$

In what follows, we focus on stationary monetary outcomes in which consumption of a buyer is time-invariant and money has constant positive value. In doing so we use the price in market two, $p_{2,t}$, as our reference price (this is without loss in generality, as we later prove). Since market one consumption depends on the size and composition of the agent's savings, these two elements must be constant in a stationary equilibrium:

$$\frac{M_{j,t}}{p_{2,t}} = \frac{M_{j,t+1}}{p_{2,t+1}} \text{ and } \frac{p_A A_{j,t}}{p_{2,t}} = \frac{p_A A_{j,t+1}}{p_{2,t+1}}.$$

At the end of $t-1$ aggregate nominal savings must equal the amount of cash available, i.e.,

$$\rho(M_{H,t} + p_A A_{H,t}) + (1-\rho)(M_{L,t} + p_A A_{L,t}) = \mathcal{M}_t$$

so that in a stationary equilibrium aggregate real balances are time-invariant and inflation equals the money stock's rate of growth:

$$\frac{\mathcal{M}_{t-1}}{p_{2,t-1}} = \frac{\mathcal{M}_t}{p_{2,t}} \Rightarrow \frac{p_{2,t}}{p_{2,t-1}} = \frac{\mathcal{M}_t}{\mathcal{M}_{t-1}} = \pi. \tag{3}$$

[3] Later it will be evident this notation is not restrictive. Sellers have identical and linear cost functions, so they produce the same quantity in market one. Since portfolios can be heterogeneous across types and buyers cannot produce, consumption may differ by type in market one. In market two everyone can produce so consumption is identical across types, and production heterogeneity in equilibrium will only correspond to heterogeneity in wealth (from idiosyncratic trading schocks) but not differences in type.

Since $\mathcal{M}_t = \mathcal{M}_{t-1}\pi$, then $\{T_t\}$ must be such that

$$\begin{array}{rl}\mathcal{M}_{t-1}\pi = & T_t + \rho\left(M_{H,t} + A_{H,t} - \frac{\alpha_H}{2}(1-p_\ell)(A_{H,t} - A_{b,H,t})\right) \\ & +(1-\rho)\left(M_{L,t} + A_{L,t} - \frac{\alpha_L}{2}(1-p_\ell)(A_{L,t} - A_{b,L,t})\right).\end{array} \tag{4}$$

That is, to sustain the deterministic money growth rate π, the per-capita money transfer in market two must equal the desired end-of-period cash supply, $\mathcal{M}_{t-1}\pi$, minus the cash available at the beginning of market two. The latter includes initial money balances $M_{j,t}$ and money associated to the redemption of bonds in market one or two, $A_{j,t} - \frac{\alpha_j}{2}(1 - p_\ell)(A_{j,t} - A_{b,j,t})$. Thus, $\{T_t\}_{t=1}^{\infty}$ is generally not a constant sequence, although it is perfectly announced since it is based on the asset holdings at the beginning of the period.

In what follows we work with real variables using $p_{2,t}$ as a normalizing factor, denoting

$$\bar{m}_t = \frac{\mathcal{M}_t}{p_{2,t}},\ \omega_{j,t} = \frac{\Omega_{j,t}}{p_{2,t}},\ m_{k,j,t} = \frac{M_{k,j,t}}{p_{2,t}},\quad a_{b,j,t} = \frac{A_{b,j,t}}{p_{2,t}},\ \tau_t = \frac{T_t}{p_{2,t}} \text{ and } p_t = \frac{p_{1,t}}{p_{2,t}}.$$

To simplify notation, we omit time subscripts and use '$'$' to denote next-period variables. Hence, real balances (2), at any date t, are

$$\begin{aligned} m_{b,j} &= m_j + p_\ell(a_j - a_{b,j}) - pc_j \\ m_{s,j} &= m_j + pc \\ m_{n,j} &= m_j. \end{aligned} \tag{5}$$

We are now ready to study the stationary equilibria. In this framework agents choose actions in order to maximize their expected discounted lifetime utility from consumption of goods in market one and two. Thus, given the recursive nature of the problem we can use a dynamic programming approach to describe the problem faced by a representative agent of type j at any date. We will let $V_j(\omega_j)$ be the expected lifetime utility of this agent when he starts the period with ω_j, before trading shocks are realized. Let $W_j(\omega_{k,j})$ be the expected lifetime utility from entering market two with $\omega_{k,j}$.

4.1 The second market

We use a functional equation to formalize the agent's problem at the start of market two. Specifically,

$$W_j(\omega_{k,j}) = \max_{q,q_k,\omega'_j \geq 0} \{U(q) - q_k + \beta_j V_j(\omega'_j)\}$$

subject to the real resource constraint

$$q + \pi(m'_j + p_A a'_j) = q_k + m_{k,j} + a_{k,j} + \tau. \tag{6}$$

The constrain holds with equality due to non-satiation. The resources available to the agent partly depend on the realization of the trading shock k, as he has $m_{k,j}$ real balances carried over from market one, and $a_{k,j}$ receipts from matured bonds. Other resources are q_k non-negative receipts from current sales of general goods and the lump-sum real-balances transfer τ. These resources can be used to finance current consumption q, to buy a'_j bonds at price p_A, or simply to carry m'_j real money balances into tomorrow's markets (short-selling is not allowed). Of course, the variable π multiplies a'_j and m'_j since these are nominal assets and nominal prices can vary across dates.

Indeed, the composition of savings will depend on the expected rates of return on cash and bonds. We emphasize that agents can save only with money or bonds and cannot lend to each other (in particular, the most patient cannot lend to the less patient) because the structure of the environment severs all future (direct and indirect) links among current trade partners.

Rewriting (6) as

$$q_k = q + \pi(m'_j + p_A a'_j) - (m_{k,j} + a_{k,j} + \tau)$$

and conjecturing $q_k \geq 0$, then we have[4]

$$W_j(\omega_{k,j}) = \max_{q,\omega'_j \geq 0} \{U(q) - q - \pi(m'_j + p_A a'_j) + m_{k,j} + a_{k,j} + \tau + \beta_j V_j(\omega'_j)\}. \quad (7)$$

A first important result emerges.

Result 1. *In a monetary equilibrium*

$$\frac{\partial W_j(m_{k,j})}{\partial m_{k,j}} = \frac{\partial W_j(\omega_{k,j})}{\partial a_{k,j}} = 1 \quad \text{for } j = H, L \quad (8)$$

The result hinges on the linearity of production disutility in market two and the use of competitive pricing (linear in the quantity sold). Since goods are sold for cash it follows that the marginal value of any asset in equilibrium must simply reflect the price of real balances, which is one. The economic implication is the marginal valuations of real balances and bonds in market two are identical and—most importantly—do not hinge on the agent's type, wealth $\omega_{k,j}$ or trade shock k.

In short, this model allows us to disentangle the agents' portfolio choices from their trading histories since

$$W_j(\omega_{k,j}) = W_j(0) + m_{k,j} + a_{k,j} \quad (9)$$

[4]Of course we must verify that $q_k \geq 0$ for all k in equilibrium. Note also that real balance transfers or taxes, τ, must also be time-invariant in a stationary equilibrium.

i.e. the agent's expected value from having $\omega_{k,j}$ at the start of market two is the expected value $W_j(0)$ from having no wealth (letting $\omega_j = (0,0) \equiv 0$) plus the current real value of wealth $m_{k,j} + a_{k,j}$. This implies agents of identical type exit market two with identical portfolios ω'_j, independent of their trading histories—much as in Lagos and Wright (2003). However, different types *might* choose different portfolios, as we demonstrate next.

Start by observing that by (7) we have $q = q^*$. That is, under the conjecture $q_k \geq 0$, then everyone can consume the same amount q^* in market two, *independent* of his asset holdings. The reason is agents in market two can produce any amount at constant marginal cost. Thus we have

$$W_j(\omega_{k,j}) = \quad U(q^*) - q^* + m_{k,j} + a_{k,j} + \max_{\omega'_j \geq 0} \left\{ -\pi(m'_j + p_A a'_j) + \beta_j V_j(\omega'_j) \right\}.$$

The central implication is the agents' lifetime utility—and the efficiency of the decentralized monetary solution—will hinge on the trades that take place in market one. Since these depend on the availability of sufficient—and sufficiently *liquid*—financial resources, then we expect that efficiency will impinge on the agents' portfolio decisions ω'_j. This is studied next.

Specifically, let $\lambda^a_j \geq 0$ and $\lambda^m_j \geq 0$ denote the Lagrange multipliers on desired real bonds and money holdings. The first order conditions stemming from the optimal portfolio choice are

$$\begin{array}{rlcll} \pi = & \beta_j \frac{\partial V_j(\omega'_j)}{\partial m'_j} + \lambda^m_j & \Rightarrow & 1 \geq \frac{\beta_j}{\pi} \times \frac{\partial V_j(\omega'_j)}{\partial m'_j} & (= \text{if } m'_j > 0) \\ p_A \pi = & \beta_j \frac{\partial V_j(\omega'_j)}{\partial a'_j} + \lambda^a_j & \Rightarrow & 1 \geq \frac{\beta_j}{\pi p_A} \times \frac{\partial V_j(\omega'_j)}{\partial a'_j} & (= \text{if } a'_j > 0). \end{array} \tag{10}$$

Recalling that one unit of real balances buys one unit of consumption, the left hand side of the expressions simply define the marginal cost of assets. The right hand sides define the expected marginal benefit from holding the asset—money or bonds—discounted according to time-preferences and inflation. The weak inequalities reflect an obvious no-arbitrage requirement: the benefit from buying any asset cannot surpass its cost.

At this point it is important to realize that an asset's expected value hinges on the asset's yield *but also* on its illiquidity, i.e. the loss from converting it into immediate cash. This is why bonds' returns must be discounted by p_A (second line), while money is not (first line). Since agents differ in their frequency of consumption shocks, it follows that the expected benefit of holding any asset will generally differ across types. To see how, we must study market one.

4.2 The first market

The expected lifetime utility of agent j who enters a period with ω_j must satisfy

$$V_j(\omega_j) = \quad \max_{c_j, c, a_{b,j}} \{\tfrac{\alpha_j}{2}[u(c_j) + W_j(\omega_{b,j})] + \tfrac{\alpha_j}{2}[-c + W_j(\omega_{s,j})]\} + (1-\alpha_j)W_j(\omega_{n,j}) \tag{11}$$

where as a buyer, he is subject to the real resource constraint

$$pc_j \leq m_j + p_\ell(a_j - a_{b,j}).$$

The agent maximizes his expected lifetime utility by choosing consumption of specialty goods $c_j \geq 0$ (as a buyer) or production $c \geq 0$ (as a seller). Traders can also choose to liquidate their bonds, which of course is a relevant choice for buyers. Thus, consumption c_j hinges on the relative price across markets, p, and on the available liquidity in the form of real money balances m_j and bonds having liquidation value $p_\ell a_j$. As bonds must be liquidated in their entirety, liquidation corresponds to $a_{b,j} = 0$ ($a_{b,j} = a_j$ otherwise). Clearly, wealthy buyers need not be constrained in their consumption, hence the weak inequality.

We start by determining the equilibrium relative price $p = \frac{p_1}{p_2}$. To do so, we study a seller's choice c to maximize his net continuation payoff, i.e.

$$\max_c -c + W_j(\omega_{s,j}).$$

The first-order condition is

$$-1 + \frac{\partial W_j(\omega_{s,j})}{\partial m_{s,j}} \frac{\partial m_{s,j}}{\partial c} = 0 \quad \Rightarrow \quad -1 + p = 0$$

where the implication follows from (5) and (8).

Result 2. *In a monetary equilibrium*

$$p = 1. \tag{12}$$

Why do we have a unit equilibrium relative price p? Recall that sellers are perfectly able to substitute market one for market two consumption. A sale in market one at price p_1 increases the cash that can be spent at price p_2 in market two. Thus, in equilibrium there cannot be arbitrage opportunities, $\frac{p_1}{p_2} = 1$. Were $\frac{p_1}{p_2} > 1$, then a market one seller would produce infinite amounts, as production generates constant (unit) marginal costs. No sale would take place if $\frac{p_1}{p_2} < 1$. Thus $p = 1$, an equilibrium condition we substitute in every expression that follows.

Now consider a buyer. Given some choice $a_{b,j}$, he selects non-negative consumption to maximize his current and continuation utility, or

$$\begin{array}{ll} \max\limits_{c_j \geq 0} & u(c_j) + W_j(\omega_{b,j}) \\ \text{s.t.} & c_j \leq m_j + p_\ell(a_j - a_{b,j}) \end{array}$$

Let $\lambda_j \geq 0$ be the Lagrange multiplier on the resource constraint (where $p = 1$). Since $u'(0) = \infty$ we have $c_j > 0$. Recall, from (5), that $m_{b,j}$ depends on c_j. Hence, the first-order necessary condition is

$$u'(c_j) + \frac{\partial W_j(\omega_{b,j})}{\partial m_{b,j}} \frac{\partial m_{b,j}}{\partial c_j} - \lambda_j = 0.$$

Since $\frac{\partial W_{b,j}}{\partial m_{b,j}} = 1$ by (8) and $\frac{\partial m_{b,j}}{\partial c_j} = -p$ by (5), then the first order condition is

$$u'(c_j) = 1 + \lambda_j. \tag{13}$$

Clearly, if $\lambda_j = 0$ then $c_j = c^*$, since $u'(c_j) = 1$. Otherwise, $c_j < c^*$. In short, it is individually optimal to consume $c_j \leq c^*$, thus

$$c_j = \min(m_j + p_\ell(a_j - a_{b,j}), c^*). \tag{14}$$

If we define $m^* = c^*$, then liquidating bonds might make sense only if $m_j < m^*$. Hence, we say buyer j is *liquidity constrained* if $m_j + p_\ell(a_j - a_{b,j}) < m^*$. Liquidation and savings are studied next.

4.3 The Marginal Value of Money and Bonds

To find the optimal portfolio of an agent, we must calculate the expected marginal value of each asset, $\frac{\partial V_j(\omega_j)}{\partial m_j}$ and $\frac{\partial V_j(\omega_j)}{\partial a_j}$. To do so use (5) and (9) in $V_j(\omega_j)$ to obtain

$$V_j(\omega_j) = \quad m_j + a_j + \tfrac{\alpha_j}{2}[u(c_j) - c_j - (a_j - a_{b,j})(1 - p_\ell)] + W_j(0) \tag{15}$$

where c_j satisfies (14).

Expression (15) tells us that the expected lifetime utility at the start of a period depends on the agent's real wealth $m_j + a_j$ and two additional elements. First, the expected utility from trade in market one. With probability $\alpha_j/2$ the agent spends c_j of his wealth on consumption and gets net utility $u(c_j) - c_j$. If the agent liquidates bonds we have $a_{b,j} = 0$ and we must account for the capital loss $a_j(1 - p_\ell)$. Second, there is the continuation payoff $W_j(0)$.

Equation (15) is useful as it makes it obvious that $c_j = c^*$ is individually optimal because it maximizes the agent's net utility in market one. Additionally, it makes it simple to calculate the equilibrium marginal value of assets. Specifically,

$$\frac{\partial V_j(\omega_j)}{\partial m_j} = \quad 1 + \frac{\alpha_j}{2}[u'(c_j) - 1]\frac{\partial c_j}{\partial m_j}$$

where $\frac{\partial c_j}{\partial m_j} = 1$ if the agent is liquidity constrained and zero otherwise (from (14)). It follows that $V_j(\omega_j)$ is strictly concave in real balances if buyer j is liquidity constrained and linear otherwise:

$$\frac{\partial V_j(\omega_j)}{\partial m_j} = \begin{cases} 1 - \frac{\alpha_j}{2}[1 - u'(c_j)] & \text{if } m_j + p_\ell(a_j - a_{b,j}) < m^* \\ 1 & \text{otherwise.} \end{cases} \tag{16}$$

For a cash-constrained buyer, the marginal value of real balances is decreasing in m_j, since $u''(c_j) < 0$.

Furthermore,

$$\frac{\partial V_j(\omega_j)}{\partial a_j} = \quad 1 + \frac{\alpha_j}{2}[u'(c_j) - 1]\frac{\partial c_j}{\partial a_j} - \frac{\alpha_j}{2}(1 - p_\ell)(1 - \frac{\partial a_{b,j}}{\partial a_j})$$

so the bond's marginal value depends on what the agent does with it. If it is used to finance market one consumption, then $\frac{\partial c_j}{\partial a_j} = p_\ell(1 - \frac{\partial a_{b,j}}{\partial a_j})$, $\frac{\partial a_{b,j}}{\partial a_j} = 0$ and $\frac{\partial c_j}{\partial a_j} = p_\ell$. Otherwise, $a_{b,j} = a_j$ so that $\frac{\partial a_{b,j}}{\partial a_j} = 1$ and $\frac{\partial c_j}{\partial a_j} = 0$. In short, $V_j(\omega_j)$ is strictly concave in a_j if a buyer is liquidity constrained, and linear otherwise:

$$\frac{\partial V_j(\omega_j)}{\partial a_j} = \begin{cases} 1 - \frac{\alpha_j}{2}[1 - p_\ell u'(c_j)] & \text{if} \quad a_{b,j} = 0 \quad \text{and} \quad m_j + p_\ell a_j < m^* \\ 1 - \frac{\alpha_j}{2}(1 - p_\ell) & \text{if} \quad a_{b,j} = 0 \quad \text{and} \quad m_j + p_\ell a_j \geq m^* \\ 1 & \text{if} \quad a_{b,j} = a_j. \end{cases} \tag{17}$$

The expression indicates that the bond's marginal value always reflects the price of real balances (which is one). If bonds are liquidated to finance consumption (first line) this value is adjusted by $-\frac{\alpha_j}{2}[1 - p_\ell u'(c_j)]$. This is the expected gain or loss from having p_ℓ additional cash ready to spend. This term is likely to be positive when cash constraints are severe (i.e., c_j is small), as there is a large marginal benefit from cashing bonds to buy consumption. Of course, if the agent is not liquidity constrained (second line), the early cashing of bonds generates a capital loss $-(1 - p_\ell)$ and no benefit. This loss is absent if bonds are not liquidated (third line).

Thus, the central observation is that illiquid bonds will be valued dissimilarly across agent types, primarily due to their heterogeneity in consumption risk, governed by α_j.

In equilibrium, this induces heterogeneity in the returns *expected* by the different agent types, as we demonstrate next.

4.4 Yields and Rates of Return

Start by considering gross nominal yields (for real, deflate by π). That on money is one while that on bonds is $1+i = \frac{1}{p_A}$, and they are both deterministic. Now consider nominal rates of return. Abstract from marginal consumption utility, for the moment. Then, the return on money is deterministic (it is the yield) and the return on illiquid bonds is deterministic only if they are held until maturity. Due to capital losses, early redemption promises an expected (nominal) rate of return which is type-dependent, $\frac{1}{p_A}[1-\frac{\alpha_j}{2}(1-p_\ell)]$ for agent j.

Of course, if assets finance market one consumption we must account for marginal consumption utility. Using (3), (10) and (16)-(17), the agents' portfolio choices in equilibrium must satisfy the following Euler equations:

$$\begin{array}{lll} 1 \geq & \frac{\beta_j}{\pi}\left\{1+\frac{\alpha_j}{2}[u'(c_j)-1]\right\} & (= \text{if } m_j > 0) \\ 1 \geq & \frac{\beta_j}{\pi p_A}\left\{1+\frac{\alpha_j}{2}[p_\ell u'(c_j)-1]\right\} & (= \text{if } a_j > 0 \text{ and } a_{b,j} = 0) \\ 1 \geq & \frac{\beta_j}{\pi p_A} & (= \text{if } a_j > 0 \text{ and } a_{b,j} = a_j) \end{array} \qquad (18)$$

with discounted real expected returns on the right hand sides and unit price on the left.

The first line refers to the choice of real balances, the second and the third lines refer to the choice of bonds under early liquidation or not. The first line tells us that, in choosing how many real balances to hold, the agent evaluates three components. The first and the second are standard: the discount factor β_j and the real yield on cash $\frac{1}{\pi}$. The third component—which is non-standard—is $\frac{1}{\pi}\frac{\alpha_j}{2}[u'(c_j)-1]$, non-negative since $u'(c_j) \geq 1$ (from (14)). It can be interpreted as the expected liquidity 'premium' from having cash available in market one; it arises because money is *needed* to conduct trades in that market. This premium is the larger the more severe is the cash constraint (the smaller is c_j) and the higher is the likelihood of a consumption shock (the higher is α_j).

A similar interpretation applies to the choice of bonds, but there are two key differences. First, bonds promise a (possibly) higher real yield $\frac{1}{\pi p_A}$. Second, a dollar worth of bonds has a smaller liquidity premium $p_\ell u'(c_j) - 1$, relative to a dollar worth of cash, if bonds are illiquid. It is this trade-off between bonds' illiquidity and superior return that will inform the agent's portfolio decision and the efficiency of equilibrium.

We can now present the following

Definition. *Given* $\{\pi, \tau, p_A, p_\ell\}$, *a stationary monetary equilibrium is a time-invariant list* *of consumption, production, real asset holdings, and relative prices* $\{c_j, c, q, q_k, m_j, a_j, p\}$ *that satisfy (1) through (18).*

At this point, some observations are in order. To start, unlike other models of money our agents are not forced to insure against consumption shocks solely with money. They can also (or solely) insure with bonds, liquidating them in market one. In any event, recalling that it is individually optimal to consume $c_j \leq c^*$, the expressions in (18) make it evident that bonds are not sold early unless the agent is cash constrained. Hence, $m_j + p_\ell a_j \leq m^*$ in an equilibrium with constrained buyers.

Of course, the portfolio composition depends on the interest rate (e.g. bonds are not superior to money if $p_A = 1$) but also on the inflation rate and the bonds' illiquidity. The first two policy parameters, p_A and π, affect returns *identically* for everyone. However, the bond's illiquidity is unlike other policy tools, as it affects agent types differently. It distorts the expected returns *dissimilarly* across types, as $1- p_\ell$ acts as a proportional tax on liquidation, whose incidence hinges on the frequency of consumption needs.

4.5 The Efficient Allocation

It is important to discuss what allocation would be selected by a planner who is subject to the same physical constraints faced by the agents, and weights each agent identically. Precisely, we must consider a planner that faces a sequence of static problems of maximizing temporary utility subject to feasibility.

Recall that each agent trades with a new set of agents in every period, goods are perishable, and at each date there is an equal number of identical buyers and sellers on each market. It follows that the efficient allocation solves a sequence of static optimization problems directed at maximizing surplus in each market. Then, we have $u'(c_H) = u'(c_L) = 1$ and $U'(q_H) = U'(q_L) = 1$ so that $c_j = c_{s,j} = c^*$ and $q_j = q_{s,j} = q^*$ is optimal at each date for $j = H, L$ (details in the Appendix).

Clearly, policy can affect economic outcomes; portfolio choices affect the real balances available in market one that in turn affects the feasible trades. One naturally wonders whether the efficient allocation can be sustained when bonds pay zero yields. The reason being, a common theoretical result indicates that selection of $i = 0$ and an appropriate deflation rate—an intervention known as the Friedman rule—is the optimal course of action. This is the first question we look into.

5 The Failure of the Friedman Rule

We start by reporting a useful result.

Result 3. *In a monetary equilibrium we must have* $\pi \geq \beta_H$.

Proof. By way of contradiction, suppose an equilibrium exists where $\pi < \beta_H$. Consider $j = H$ in the first line of (18). For $m_j > 0$ we would need $\pi \geq \beta_H + \beta_H \frac{\alpha_H}{2}[u'(c_H) - 1] \geq \beta_H$. This is in contradiction with the conjecture $\pi < \beta_H$. Thus $\pi \geq \beta_H$.■

The result that an excessive rate of return on money is inconsistent with monetary equilibrium is an obvious no-arbitrage result. It is in line with the finding of Becker (1980) for an economy with a fixed stock of capital, whose equilibrium rate of return cannot exceed the lowest rate of time preference. Intuitively, in a monetary economy money's value cannot grow too fast or agents would not spend it. The key observation here is money's value cannot grow at a rate $\frac{1}{\pi}$ that is superior to the return desired—so to speak—by the most patient agent. Setting $\pi < \beta_H$ creates an arbitrage opportunity for the most patient agents, as the return on money exceeds their shadow interest rate $\frac{1}{\beta_H}$. This of course is inconsistent with equilibrium.

The implication is policy makers are constrained in their ability to give cash a return that is *sufficiently* attractive for everyone. Thus, inefficiencies are to be expected when saving can only take the form of cash. To formalize this intuition we remove the incentives to save with bonds by setting $i = 0$, as Friedman suggested.[5] Then we ask the question: is there any $\pi \geq \beta_H$ that sustains the efficient allocation?

Result 4. *Consider $i = 0$ and $\pi \geq \beta_H$. A unique monetary equilibrium exists and money holdings are heterogeneous, $m^* \geq m_H > m_L > 0$. The allocation is inefficient.*

Proof. Let $p_A = 1$ so $i = 0$. From (18) we get:

$$\begin{array}{lll} \pi \geq & \beta_j \left\{1 + \frac{\alpha_j}{2}[u'(c_j) - 1]\right\} & (= \text{if } m_j > 0) \\ \pi \geq & \beta_j \left\{1 + \frac{\alpha_j}{2}[u'(c_j)p_\ell - 1]\right\} & (= \text{if } a_j > 0 \text{ and } a_{b,j} = 0) \end{array} \tag{19}$$

It is obvious that bonds and money are equivalent assets only if $p_\ell = 1$ (they are inferior otherwise). Thus, suppose $p_\ell = 1$ and discuss money.

To prove the equilibrium is inefficient note $\pi \geq \beta_H$ is necessary. From (18), $m_H > 0$ if

$$\pi = \quad \beta_H \left\{1 + \frac{\alpha_H}{2}[u'(c_H) - 1]\right\} \geq \beta_H \quad .$$

[5] One way to interpret this suggestion is to realize that—from a social efficiency standpoint— marginal social benefits and costs of money should match. Since the private cost of holding an extra dollar is the nominal interest rate, and money is costlessly produced, we should set $i = 0$. As real interest rates are positive, Friedman suggested a deflation equal to the real interest rate, i.e. the unique discount factor in a representative agent model. In our model we have more than one discount factor, but this is irrelevant, since we have established $\pi = \beta_H$ is the best return money can ever give.

If $\pi > \beta_H$ then $c_H < c^*$ and $m_H < m^*$. If $\pi = \beta_H$ then $c_H = c^*$ and $m_H = m^*$. Thus, suppose $\pi = \beta_H$. Now, $m_L > 0$ requires

$$\pi = \quad \beta_L \left\{1 + \tfrac{\alpha_L}{2}[u'(c_L) - 1]\right\} = \beta_H.$$

Since $\beta_L < \beta_H$, it follows that $c_L < c^*$ and $m_L < m^*$. Hence, if $p_A = 1$ then a unique stationary monetary equilibrium exists in which $m^* \geq m_H > m_L > 0$ and $c^* \geq c_H > c_L > 0$. In equilibrium $\lim_{\pi \to^+ \beta_H} m_H = m^*$ so $\lim_{\pi \to^+ \beta_H} c_H = c^*$; also, $\frac{\partial c_L}{\partial \pi} < 0$. Thus, the Friedman rule is a second best.■

What is the intuition? When $i = 0$ effectively we have a model where agents insure against consumption shocks with money. Due to discounting disparities equilibrium returns must obey the no-arbitrage restriction $\pi \geq \beta_H$, so the more impatient will tend to under-insure. This leaves them liquidity constrained in market one, which creates an inefficiency. Of course, setting $\pi = \beta_H$ leads to a second best (since $m_H = m^*$).

This result seems quite robust. The Friedman rule should fail to achieve the first best in any model in which money has an explicit transactions role and agents 'price' unequally future consumption. In fact, lowering the return on bonds to that of money (by setting $i = 0$) seems to be the *source* of the problem. It eliminates the opportunity cost of holding money (which is good) but it fails to provide adequate incentives for everyone to save enough (which is bad), since $\pi \geq \beta_H$. Thus, we next consider a policy where $i > 0$. Before doing so, however, several remarks are in order.

The Friedman rule does not fail to be a first best simply because bonds are illiquid. Setting $p_\ell = 1$ under $i = 0$ simply makes money and bonds indistinguishable financial instruments. Also, the result does not hinge on the mere existence of some arbitrary heterogeneity element that gives different agents incentives to hold unequal money balances. In fact, the Friedman rule can be quite effective in eliminating equilibrium heterogeneity in real-balances.

To see why, consider $\beta_H = \beta_L = \beta$, while retaining the assumption of disparities in trade shocks, $\alpha_H > \alpha_L$. Set $i = 0$ so, from (19), a unique monetary equilibrium exists for $\pi > \beta$. Specifically, we have

$$\pi = \quad \beta \left\{1 + \tfrac{\alpha_H}{2}[u'(c_H) - 1]\right\} = \quad \beta \left\{1 + \tfrac{\alpha_L}{2}[u'(c_L) - 1]\right\}.$$

For $\pi > \beta$ balances and consumption are heterogeneous, $c^* > c_H > c_L$ and $m^* > m_H > m_L$. Types L under-insure as they do not need cash as frequently as types H (the opposite occurs if $\alpha_H < \alpha_L$). As $\pi \to^+ \beta$ real balances all converge to m^* as agents become indifferent between having a dollar today or one tomorrow.[6] In this case, trade-frequency considerations do not enter saving decisions (see also Boel and Camera, 2004).

6 Using Bonds to Finance Consumption

We now want to demonstrate that the efficient allocation can be sustained when the bonds' yield is positive. To simplify our task, we start by proving that such an allocation is inconsistent with agents holding mixed portfolios.

Result 5. *Consider $i > 0$. If in equilibrium $m_j > 0$ then $c_j < c^*$.*

Proof. Let $i > 0$. We want to show that an agent who holds bonds and money in equilibrium must be liquidity constrained.

1. We start by proving that $m_j < m^*$. Let $a_j \geq 0$. By way of contradiction, suppose $m_j \geq m^*$. Here $c_j = c^*$ and bonds are not liquidated. From (18), we need $\pi = \beta_j$ for $m_j > 0$, which implies $\pi < \frac{\beta_j}{p_A}$. This is inconsistent with equilibrium as agents would buy infinite bonds. Thus, it must be that $m_j < m^*$ so consider $0 \leq m_j < m^*$.

2. If $m_j = 0$, then for $c_j > 0$ we need $a_j > 0$ and $a_{b,j} = 0$. From (18), we need $\pi = \frac{\beta_j}{p_A}\left[1 - \frac{\alpha_j}{2}(1 - p_\ell u'(c_j))\right]$. Notice that if $-\frac{\alpha_j}{2}(1 - p_\ell u'(c_j)) < 0$ then $\pi < \frac{\beta_j}{p_A}$. This, however, is *not* inconsistent with equilibrium since fractions of bonds cannot be liquidated. Thus, the agent would not buy infinite amounts of bonds and avoid liquidating them as (given $m_j = 0$) his marginal utility of consumption would be infinite. Maximization requires $u'(c_j) \geq 1$ so the agent will not buy more bonds than necessary to acquire c^*. Hence $p_\ell a_j \leq c^*$.

3. Consider $0 < m_j < m^*$ and suppose $a_j > 0$. Then, using (18), the following conditions must hold in equilibrium:

$$\pi = \begin{cases} \beta_j \left\{1 + \frac{\alpha_j}{2}[u'(c_j) - 1]\right\} & \text{for} \quad m_j > 0 \\ \frac{\beta_j}{p_A}\left\{1 + \frac{\alpha_j}{2}[u'(c_j)p_\ell - 1]\right\} & \text{for} \quad a_j > a_{b,j} = 0 \\ \frac{\beta_j}{p_A} & \text{for} \quad a_j = a_{b,j} > 0 \end{cases} \tag{20}$$

   The first line in (20) is an equality since we are conjecturing $m_j > 0$. One of the other two lines must also hold with equality, since $a_j > 0$. The second line holds with equality, and the third with inequality, if the agents uses bonds to finance first

[6] For $\pi = \beta$ a continuum of monetary equilibria exists. The reason is price indeterminacy, as any sequence $\{p_t\}$ which is consistent with $p_{t+1}/p_t = \beta$ is an equilibrium.

market consumption. The reverse is true if bonds are held until maturity. Either way, using the first and third line of (20) we need

$$\frac{1}{p_A} - 1 \leq \frac{\alpha_j}{2}[u'(c_j) - 1]. \tag{21}$$

This no-arbitrage condition, says the net interest rate can never be so high to surpass the expected marginal value of the 'liquidity services' provided by bonds, $\frac{\alpha_j}{2}[u'(c_j) - 1]$. This condition must always hold when the agent saves money. Now there are two separate cases

(i) If (21) is an equality, then $c_j < c^*$ since $i > 0$. This is independent of whether the second line in (20) holds with a strict inequality or not. Thus, in this case the agent has bonds and money, and he is constrained in market one.

(ii) If (21) is an inequality, then $\pi = \beta_j \left\{1 + \frac{\alpha_j}{2}[u'(c_j) - 1]\right\} \geq \frac{\beta_j}{p_A}$. Because we are conjecturing $a_j > 0$ then the second line in (20) must hold with equality:

$$\pi = \quad \beta_j \left\{1 + \tfrac{\alpha_j}{2}[u'(c_j) - 1]\right\} = \tfrac{\beta_j}{p_A} \left\{1 + \tfrac{\alpha_j}{2}[u'(c_j)p_\ell - 1]\right\} \tag{22}$$

Clearly we also need

$$\tfrac{\beta_j}{p_A} \left\{1 + \tfrac{\alpha_j}{2}[u'(c_j)p_\ell - 1]\right\} \geq \tfrac{\beta_j}{p_A} \quad \Rightarrow \quad p_\ell \geq \tfrac{1}{u'(c_j)}. \tag{23}$$

This implies $c_j < c^*$. To see why notice that $p_\ell \leq 1$ by definition, so we need $u'(c_j) \geq 1$, hence $c_j \leq c^*$. To see why $c_j \neq c^*$ note that if $c_j = c^*$ then we need $p_\ell = 1$. But then (22) cannot hold since we would have $\pi = \beta_j < \frac{\beta_j}{p_A}$.

We conclude that if $m_j, a_j > 0$ then $c_j < c^*$. This is true whether $a_{b,j} = 0$ or not. Clearly, if $a_{b,j} = 0$ then (23) must hold. The most stringent case corresponds to the smallest $u'(c_j)$, which is $u'(c_j) = 1 + \frac{2i}{\alpha_j}$, from (21). Thus (23) becomes $p_\ell \geq \frac{\alpha_j}{\alpha_j + 2i}$. Bonds cannot be too illiquid if they are used to finance market one consumption.■

In short, when bonds pay a positive yield agents who hold money and bonds must be liquidity constrained. Again, this is an arbitrage argument. In fact, suppose the agent is unconstrained in his consumption in equilibrium. If the agent has money and bonds their expected rates of returns must be identical. But then the agent could accumulate wealth by holding bonds until maturity, which cannot be a stationary equilibrium.

This result suggests that perhaps the optimal policy should encourage agents to save with bonds and not money. The government could make cash an unattractive asset—for saving purposes—by selecting a sufficiently high π. Then, perhaps agents would fully insure against consumption shocks using bonds, liquidating them when needed. In the words of Tobin

> "Why not hold transactions balances in assets with higher yields than cash, shifting into cash only at the time an outlay must be made?" (1956, p.241)

The problem with this is that the most patient agents might want to buy large amounts of bonds. To see why, notice that if $c_j = c^*$ then (18) implies

$$\begin{array}{lll} \pi \geq & \beta_j & (= \text{if } m_j > 0) \\ \pi \geq & \frac{1}{p_A}\beta_j \left[1 - \frac{\alpha_j}{2}(1-p_\ell)\right] & (= \text{if } a_j > 0 \text{ and } a_{b,j} = 0) \end{array} \tag{24}$$

so that $\frac{1}{p_A}\beta_H \left[1 - \frac{\alpha_H}{2}(1-p_\ell)\right] > \beta_H \geq \frac{1}{p_A}\beta_L \left[1 - \frac{\alpha_L}{2}(1-p_\ell)\right]$. Our next objective is to prove that, in certain economies, such arbitrage opportunities can be avoided in a simple way: by making bonds sufficiently illiquid.

6.1 The Optimal Illiquidity of Bonds

We start by defining the following condition:

$$\frac{\beta_L}{\beta_H} > \frac{2-\alpha_H}{2-\alpha_L}. \tag{25}$$

Since $\alpha_H > \alpha_L$ then $\frac{2-\alpha_H}{2-\alpha_L} < 1$. Thus (25) simply limits the extent of disparities in individual discount factors. We then proceed by demonstrating that, under this condition, the efficient allocation can be achieved if bonds are sufficiently illiquid.

Result 6. *Let condition (25) be satisfied. If*

$$\begin{array}{ll} \pi > & \beta_H \\ p_\ell = & 1 - \frac{2(\beta_H - \beta_L)}{\alpha_H\beta_H - \alpha_L\beta_L} \\ p_A = & \frac{\beta_H}{\pi}\left[1 - \frac{\alpha_H}{2}(1-p_\ell)\right] \end{array} \tag{26}$$

then $c_j = c^*$ is a stationary monetary equilibrium. Here $p_A, p_\ell \in (0,1)$.

Proof. Conjecture $c_j = c^*$. Applying our previous results we must insure that agents save only with bonds and that bonds pay a positive yield. That is, we need $m_j = 0$, which requires $\pi > \beta_H$ (see the first expression in (24)). We also need $i > 0$ and $a_j > a_{b,j} = 0$ (since the agent holds no money). Thus, focus on the second expression in (24), which must hold with equality for all j

$$\pi = \frac{\beta_j}{p_A}\left[1 - \frac{\alpha_j}{2}(1-p_\ell)\right]. \tag{27}$$

Consider $j = H$. Then we need

$$p_A = \frac{1}{\pi}\beta_H \left[1 - \frac{\alpha_H}{2}(1 - p_\ell)\right] \equiv \frac{1}{\pi}h(p_\ell) \tag{28}$$

which defines uniquely p_A as a function of π. Since $\pi > \beta_H$ then $p_A < 1 - \frac{\alpha_H}{2}(1 - p_\ell)$. Thus nominal interest rates are bounded strictly away from zero.

Now consider $j = L$. Equation (27) for $j = L$ and $p_A = \frac{h(p_\ell)}{\pi}$ implies

$$\beta_H \left[1 - \tfrac{\alpha_H}{2}(1 - p_\ell)\right] = \beta_L \left[1 - \tfrac{\alpha_L}{2}(1 - p_\ell)\right] \quad \Rightarrow \quad \tfrac{\beta_L}{\beta_H} = \tfrac{2-\alpha_H(1-p_\ell)}{2-\alpha_L(1-p_\ell)}.$$

Hence, there exists a unique

$$p_\ell = 1 - \frac{2(\beta_H - \beta_L)}{\alpha_H\beta_H - \alpha_L\beta_L} \tag{29}$$

such that (27) holds for all j. We have $p_\ell > 0$ only if (25) holds, a condition we retain. Since $\beta_H > \beta_L$ and $\alpha_H > \alpha_L$ then $p_\ell < 1$. However, $p_\ell = 1$ if $\beta_H = \beta_L$. Note also that $p_\ell > p_A$ if π is large.

In equilibrium $a_j = a$ for $j = L, H$. Since $p_A a = \bar{m}$ and $a' = a$, then we have $a' = a = \frac{\bar{m}}{p_A}$. Taxes are

$$\tau = \bar{m} - \frac{\bar{m}}{p_A}\left[1 - (1 - p_\ell)\left(\rho\frac{\alpha_H}{2} + (1 - \rho)\frac{\alpha_L}{2}\right)\right]$$

i.e. cash at the end of the period, $\bar{m}$, minus the payments to bond holders, $\frac{\bar{m}}{p_A}$, adjusted for the liquidation cost, $\bar{m}\frac{1-p_\ell}{p_A}\left(\rho\frac{\alpha_H}{2} + (1 - \rho)\frac{\alpha_L}{2}\right)$. Finally, it can be proved that $q_k \geq 0$ if $U'(x)$ is sufficiently larger than $u'(x)$ for $x \in \mathbb{R}_+$ (see the Appendix).■

In short, when the most patient agents are also those who are more prone to consumption shocks, then two elements are needed to sustain the efficient allocation: savings with bonds must be encouraged, by setting $i > 0$ and setting $\pi > \beta_H$, and bonds must be illiquid, $p_\ell < 1$. What is the intuition? First, we know that deflation cannot be too pronounced in a monetary equilibrium, therefore the impatient agents would under-insure by using cash. Consequently, we must give bonds a return superior to cash.

However, the patient agents would demand infinite quantities of bonds if $p_\ell = 1$. Thus, we need to lower the return on bonds for these agents. Since types H need cash more frequently than agents L, this can be done by setting $p_\ell < 1$. When (25) holds, a unique $p_\ell \in (0, 1)$ exists that equates the present values of returns across agents:

$$\beta_H \left[1 - \frac{\alpha_H}{2}(1 - p_\ell)\right] = \beta_L \left[1 - \frac{\alpha_L}{2}(1 - p_\ell)\right].$$

The equilibrium p_ℓ falls as discounting disparities increase, which is why heterogeneity in discounting cannot be too extreme, i.e. (25).

By substituting (29) into (28), we obtain the nominal interest rate that sustains the efficient equilibrium:

$$1+i=\frac{\pi}{\beta_H}\theta \quad \text{where} \quad \theta=\frac{\alpha_H\frac{\beta_H}{\beta_L}-\alpha_L}{\alpha_H-\alpha_L}\geq 1.$$

Nominal interest rates are a function of a weighted measure of the agents' discount factors, θ, with weights given by the frequencies of consumption shocks.

We see that the model is consistent with the notion of existence of a 'Fisher effect,' as i fully accounts for inflationary pressure, rising or falling, but leaving the allocation unaffected. In particular, bonds dominate cash in rate of return, which is why no one saves with cash. Bond yields also include a liquidity premium, captured by θ, since an increase in the bonds' illiquidity lessens their attractiveness. In environments where the efficient equilibrium is associated to a lower p_ℓ (hence a higher θ), we see that the bonds' yield must be higher. As discounting disparities vanish, so does the need for illiquidity and

$$\lim_{\beta_H,\beta_L\to\beta} 1+i=\frac{\pi}{\beta}$$

i.e. the optimal real yield converges to the (common) rate of time preference, $\frac{1}{\beta}$.

7 Is the Optimal Policy Necessarily Deflationary?

We have seen that the availability of nominal bonds allows to sustain the efficient outcome even when monetary policy is inflationary. In short, the Friedman rule is not uniquely optimal and sometimes it is simply suboptimal. To build intuition consider the case $\beta_H=\beta_L=\beta$ and suppose with have both money and bonds.

There are two ways to sustain $c_j=c^*$ for all j in this economy. Both hinge on the availability of some asset that can be easily transformed into consumption, and that offers a real yield $\frac{1}{\beta}$. A first possibility is to induce agents to save with cash, insuring that cash pays the return $\frac{1}{\beta}$. This is done by *lowering* the yield on bonds to that of money. Since the real yield on cash is $\frac{1}{\pi}$, and on bonds is $\frac{1+i}{\pi}$, we must set $i=0$. Since cash cannot pay interest and cannot be bought at a discount, then a deflation must be run at rate $\pi=\beta$. Here, we are at the Friedman rule and money and bonds are perfect substitutes if $p_\ell=1$.

Alternatively, the government can give incentives to save with interest-paying bonds that can be easily redeemed for cash. This is equivalent to setting $\pi>\beta$ while selling bonds at price $p_A=\frac{\beta}{\pi}$, standing ready to costlessly redeem them, i.e. setting $p_\ell=1$. Here, agents save with bonds—not money—and obtain the real return $\frac{1}{\beta}$, independent of

π. In short, deflations in this environment are unnecessary for efficiency, as long as assets exist that can be easily liquidated when a consumption opportunity arises.

In the absence of a 'cash management' technology that allows for cheap liquidation of bonds, then one can set

$$p_A = \frac{\beta}{\pi}[1 - \frac{1-p_\ell}{2}\max(\alpha_H, \alpha_L)]$$

and then engage in rationing bonds' purchases. This is reminiscent of the market for U.S. government EE series savings bonds, which are registered illiquid bonds that cannot be purchased in quantities that exceed a fixed nominal amount (currently \$60,000). We note that this same rationing strategy would sustain an efficient allocation when $\beta_H > \beta_L$ but $\alpha_H \leq \alpha_L$.

8 Final Remarks

The analysis in this paper offers us two basic lessons. A first lesson is that heterogeneity in discounting blunts the appeal of the Friedman rule, due to equilibrium heterogeneity in desired real balances.

Under zero interest rates, agents essentially must rely on the available stock of fiat money as a means to insure against consumption risk. A simple arbitrage argument indicates that a deflationary policy cannot achieve the first best. The reason is cash cannot promise a return greater than the discount factor of the most patient agents, much as it happens for the return on capital in Becker (1980). Therefore, the more impatient will under-insure, which is detrimental to efficiency.

Under-insurance means that in equilibrium agents hold different amounts of the available stock of nominal assets although, unlike Becker (1980), *everyone* holds some. These findings should obtain in any environment where money is essential to execute trades, since it is simply an arbitrage argument.

A second lesson is that bonds should provide a positive yield in order to sustain an efficient equilibrium. Furthermore, under certain conditions, an additional friction is needed: bonds should be *illiquid*, i.e. should be convertible into immediate consumption less efficiently than cash.

In the model, this necessity stems from difference in desired rates of return and in consumption needs. Illiquidity is a friction that removes arbitrage opportunities if the individuals who have the lowest discount rate are also those who are more severely exposed to consumption risk. Although this result is less general, it suggests one (more) reason as to why illiquid government bonds might be desirable financial instruments.

References

Aliprantis R., Camera G. and Puzzello D. (2004). A Random Matching Theory. Krannert Working Paper # 1168, Purdue University.

Becker R.A. (1980). On the Long-Run Steady State in a Simple Dynamic Model of Equilibrium with Heterogeneous Households. *Quarterly Journal of Economics*, 365-382

Berentsen A., and C. Strub (2004). On the Friedman Rule with Heterogeneous Agents. Mimeo, University of Basel.

Berentsen A, G. Camera and C. Waller, (2003). The Distribution of Money Balances and the Non-Neutrality of Money. *Manuscript*, Purdue University.

Boel P. and Camera G. (2004). Heterogeneity and the (sub) optimality of the Friedman rule. *Manuscript*, Purdue University.

Camera G., and Corbae D. (1999). Money and Price Dispersion. *International Economic Review,* 40, 985-1008.

Chari V.V., Christiano L. and Kehoe P. (1996,) Optimality of the Friedman Rule in Economies with Distorting Taxes, *Journal of Monetary Economics*, 37, 203-223.

Correia, I., and P. Teles, 1996. Is the Friedman rule optimal when money is an intermediate good?. *Journal of Monetary Economics*, 38, 223-244.

da Costa, C., and I. Werning, 2003, On the Optimality of the Friedman Rule with Heterogeneous Agents and Non-Linear Income Taxation, *Manuscript*, MIT.

Faig, M.,1988, Characterization of the Optimal Tax on Money when it functions as a Medium of Exchange, *Journal of Monetary Economics* 22, 137-148.

Freeman S. (1993). Resolving Differences over the Optimal Quantity of Money. *Journal of Money, Credit and Banking*, 25, 801-811.

Kimbrough K.P. (1986). The Optimum Quantity of Money Rule in the Theory of Public Finance. *Journal of Monetary Economics*, 18, 277-284.

Lagos R. and Wright R. (2002). A Unified Framework for Monetary Theory and Policy Evaluation." *Working pape*r, University of Pennsylvania.

Lucas R. (1980). Equilibrium in a Pure Currency Economy. In *Models of Monetary Economies*, J. Kareken and N. Wallace editors, 131-145.

Mulligan, C.B., and X. Sala-i-Martin (1997). The Optimum Quantity of Money: Theory and Evidence. *Journal of Money, Credit and Banking*, 29, 687-715.

Paal, B. and Bruce D, Smith (2000). The Suboptimality of the Friedman Rule and the Optimum Quantity of Money. *Manuscript*, Stanford University and University of Texas, Austin.

Phelps E. S. (1976). Inflation in the Theory of Public Finance. *Scandinavian Journal of Economics,* 75, 67-82.

Shi, S., (1997). A Divisible Search Model of Fiat Money. *Econometrica,* 65, 75-102.

Shi S. (2003). Nominal Bonds and Interest Rates: Liquidity and Welfare. *Manuscript*, University of Toronto.

Tobin J. (1956). The Interest-Elasticity of Transactions Demand for Cash. *Review of Economics and Statistics*, 38(3), 241-247.

Townsend R. (1980). Models of Money with Spatially Separated Agents. In *Models of Monetary Economies*, J. Kareken and N. Wallace editors, 265-303.

Williamson S. D. (1996). Sequential Markets and the Suboptimality of the Friedman Rule. *Journal of Monetary Economics*, 37, 549-572.

Williamson S. D. (2004). Limited Participation, Private Money, and Credit in a Spatial Model of Money. *Economic Theory*, 24(4), 857-875.

Woodford M. (1990). The Optimum Quantity of Money. In *Handbook of Monetary Economics.*

Appendix

Proof of $q_k \geq 0$

We now want to provide conditions that guarantee $q_k \geq 0$ in the efficient equilibrium described by Result 6. From Result 2 we have $p_1 = p_2 = p$, and we know $q = q^*$. These results and the budget constraint (6) imply

$$q_k = q^* + \pi(m'_j + p_A a'_j) - (m_{k,j} + a_{k,j} + \tau)$$

In the efficient equilibrium agents save only with bonds, $m'_j = 0$, and $a_j = a = a'$ for all j. Since $p_A a = \bar{m}$ we have $a' = a = \frac{\bar{m}}{p_A}$. Thus

$$q_k = q^* + \bar{m}\pi - (m_{k,j} + a_{k,j} + \tau)$$

From now on we are going to focus on the seller's case, since $q_b > q_s$. We have $m_{s,j} = c^* = \frac{\bar{m}p_\ell}{p_A}$ and $a_{s,j} = a = \frac{\bar{m}}{p_A}$. Therefore,

$$\begin{aligned} q_s &= q^* + \bar{m}\pi - c^* - \tfrac{\bar{m}}{p_A} - \tau \\ &= q^* + \tfrac{c^* p_A \pi}{p_\ell} - c^* - \tfrac{c^*}{p_\ell} - c^* \tfrac{p_A}{\bar{m}p_\ell}\tau \\ &= q^* - c^* + \tfrac{c^* p_A \pi}{p_\ell} - \tfrac{c^* p_A}{p_\ell} - c^* \tfrac{1-p_\ell}{p_\ell}\left(\rho\tfrac{\alpha_H}{2} + (1-\rho)\tfrac{\alpha_L}{2}\right) \end{aligned}$$

since

$$\tau = \bar{m} - \tfrac{\bar{m}}{p_A} + \bar{m}\tfrac{1-p_\ell}{p_A}\left[\rho\tfrac{\alpha_H}{2} + (1-\rho)\tfrac{\alpha_L}{2}\right].$$

Therefore

$$q_s \quad = q^* - c^*\left(1 + (1-\pi)\tfrac{p_A}{p_\ell} + \tfrac{1-p_\ell}{p_\ell}\left(\rho\tfrac{\alpha_H}{2} + (1-\rho)\tfrac{\alpha_L}{2}\right)\right) \tag{30}$$

Since the term multiplying c^* is greater than one, then in order to have $q_s > 0$, q^* must be sufficiently larger than c^*. Since in the efficient equilibrium $U'(q^*) = u'(c^*) = 1$, then (30) implies we need $U'(x) > u'(x)$, i.e. the marginal utility of consumption in market two must be sufficiently higher than the marginal utility of consumption in market one for any amount of consumption $x \in \mathbb{R}_+$.

The Planner's Problem

At each date the planner has buyers and sellers of identical mass $\frac{1}{2}[\rho\alpha_H + (1-\rho)\alpha_L]$, in market one, and of mass one in market two. All buyers have identical preferences and all sellers have identical unit marginal production cost. Assuming the planner weights agents identically it is obvious that the sellers' marginal utilities must be identically equal to one. Production can be assigned to sellers differently—due to fixed marginal production costs—and equal production is a possibility. To see it, notice that in each period the planner solves

$$\max_{c_j, c_j^s, q_j, q_j^s} \frac{1}{2} \{\rho\alpha_H[u(c_H) - c_{s,H}] + (1-\rho)\alpha_L[u(c_L) - c_{s,L}]\} + \rho[U(q_H) - q_{s,H}] + (1-\rho)[U(q_L) - q_{s,L}]$$

subject to

$$\rho\alpha_H c_H + (1-\rho)\alpha_L c_L = \rho\alpha_H c_{s,H} + (1-\rho)\alpha_L c_{s,L}$$

$$\rho q_H + (1-\rho) q_L = \rho q_{s,H} + (1-\rho) q_{s,L}.$$

Here $c_j = c_{s,j} = c^*$ and $q_j = q_{s,j} = q^*$ maximizes trade surplus in market one, $u(c_j) - c_{s,j}$, and in market two, $U(q_j) - q_{s,j}$.