Transactions and Mechanism Design*

Thorsten Koeppl
Queen's University
Kingston, Ontario
thor@qed.econ.queensu.ca

Cyril Monnet
European Central Bank
Frankfurt, Germany
cyril.monnet@ecb.int

Ted Temzelides
University of Pittsburgh
Pittsburgh, PA 15260
tedt@pitt.edu

Preliminary draft. First version: July 2004

August 3, 2005

*We thank audiences at the Bank of Portugal, CEMFI, the Cleveland Fed, the European Central Bank, the SAET meetings in Vigo, Spain, the SED meetings in Budapest, Hungary, and the Stanford Institute for Theoretical Economics. We especially thank Ed Green, Benjamin Lester, and Neil Wallace for several comments and suggestions. The views expressed in the paper are not necessarily those of the ECB.

Abstract

We introduce a dynamic model in which the ability of agents to perform certain welfare improving transactions is subject to random and unobservable shocks. We characterize the optimal setup for incentive compatible transactions, the optimal *payment system*. Implementation involves assigning *balances* to individual agents and optimally adjusting these balances given the agents' histories of transactions. The existence of an equilibrium in which agents transact through a payment system requires certain caps on short-term borrowing. In the absence of *settlement*, incentive constraints imply that the full information first-best allocation cannot be supported. We introduce a periodic pattern in which several rounds of bilateral transactions are followed by centralized settlement. We consider both the case where there is discounting between transaction rounds, and the case where there is none. The first-best is supportable if settlement takes place with a sufficiently high frequency. The Friedman rule is not necessary for efficiency since a policy that leads to a constant supply of nominal balances implements the first-best. As in Levine (1991), if settlement rounds are infrequent, the first-best allocation is not supportable within a payment system that operates under a Friedman-like rule.

1 Introduction

One of the features of the economy that the Walrasian model abstracts from is the mechanism through which transactions for goods and services take place, the *payment system*. While for the study of certain questions this abstraction is one of the main strengths of the Walrasian model, this feature also makes it an inappropriate tool for the study of questions related to transactions. Thus, new models are needed to study payments, and our goal is to develop such models using mechanism design.[1]

There are several questions about the optimal structure of payment systems that motivate our work. For example, should there be binding limits, or "caps," imposed on the short-term borrowing by participants? What are the effects of reputation through repeated interactions with the system? What is the role of private information and imperfect monitoring in answering the above questions? Is settlement welfare improving? What are the features of optimal payments? Our approach emphasizes the role of private information. We are motivated by recent work by Kocherlakota (2004), who extends the model of Mirrlees (1971) and studies optimal taxation under information frictions. The design of optimal payment systems (and, more generally, of monetary policy) under imperfect monitoring is subject to a similar private information problem since participants' ability to raise liquidity is not directly observable. Importantly, some of the questions that optimal payment system design poses are inherently *dynamic* and, therefore, very hard or impossible to study within the existing literature, which is almost exclusively static.[2]

We model the participants of the system as agents who face random needs for liquidity and random opportunities to build balances that can be used to meet these needs. To perform either of these activities, they need to interact through a network consisting of a large number of participants. We employ a version of the model of exchange developed by Kiyotaki and Wright (1989,1993). This model is appropriate for our study for several reasons. First, it offers a setup in which transactions are explicit, and where it is nat-

[1] To get some idea of the magnitudes in actual systems involving bank payments, the value of the transactions processed through TARGET, the main public payment system in Europe, during March 2004 was over 40 € billions, with a daily average of about 1.7 € billions. In the United States, the average daily value of transfers through FEDWIRE, the US equivalent of TARGET, during the first quarter of 2004 was 1,683,265 $ millions.

[2] See Kahn and Roberds (1998) for a well-cited paper in this literature.

ural to study the implications of lack of commitment, imperfect monitoring, and reputation. Second, the random matching shocks that agents are subject to in this model are a tractable way of modelling random needs for liquidity. They simply capture the fact that certain participants need to transfer resources to other participants, and that these needs are subject to randomness. Third, the model is consistent with the fact that actual transactions *are* bilateral and, often, subject to private information problems. Finally, this setup naturally lends itself to mechanism design. Unlike the standard monetary theory approach, our model involves *no currency*.

Our approach is related to the dynamic contracting literature.[3] Agents in our model participate in two kinds of transactions. Some transactions can be perfectly monitored. On the other hand, other transactions are subject to imperfect monitoring. During such transactions, the payment system needs to induce truthful revelation by one of the participants. This is accomplished by exploring intertemporal incentives through adjustments in agents' balances. These play a similar role to promised utility adjustments in Green's model.[4]

A positive volume of transactions requires that certain caps are imposed on short-term borrowing. Put differently, in order for liquidity to be of value, it needs to be sufficiently scarce. Importantly, the introduction of caps implies a welfare loss as it rules out certain welfare improving transactions that would take place under full information. This is an example of the common trade-off between efficiency and incentives since truthful revelation comes at some cost. To avoid an unnecessary loss of certain transactions, caps should be set at the maximum value consistent with incentive compatibility. The existence of caps implies that the first-best allocation is not implementable in this framework. This is because a positive fraction of agents reach the cap in any given period.

We then impose a periodic pattern in which several rounds of bilateral transactions are followed by centralized settlement. We consider both the case where there is positive discounting between transaction rounds, and the

[3] See, for example, Green (1987). Other classic references include Spear and Srivastava (1987) and Atkeson and Lucas (1993).

[4] One interpretation of this setup is that agents are connected to different networks. With some probability, agents need to transact within their network, in which case the ability to perform the transaction is perfectly observable. Alternatively, an agent might transact with an agent belonging to a different network. For this to take place, the system has to ensure that the desired transaction is incentive compatible for both participants.

case where there is none.[5] In order to simplify the contracting environment, we assume linear utility in the settlement stage.[6] Unlike the standard setup in the dynamic contracting literature, agents in our model transact bilaterally. Thus, in general, payoffs depend on the agents' histories as well as on the histories of their current and future trading partners. Keeping track of such histories poses a problem that is hard to study analytically. Conveniently, the settlement stage allows us to circumvent the "distribution problem" that often limits the tractability of monetary models. We emphasize that our approach does not trivialize the distribution of agents' balances. Rather, following a mechanism design instead of a strategic approach allows us to find sufficient conditions for supporting the first best *independent* of the distribution.[7]

We find that the first-best is supportable if settlement takes place with a sufficiently high frequency. As in Levine (1991), if settlement rounds are infrequent, we demonstrate that the first-best allocation is not supportable within a payment system that operates under a Friedman-like rule. In contrast, a payment system policy that leads to constant supply of nominal balances is shown to be sufficient for implementing the first-best.

The paper proceeds as follows. Section 2 describes the basic economic environment subject to indivisibility restrictions which we subsequently drop. The main topic of the paper, concerning optimal settlement under private information, is studied in Section 3. A brief conclusion follows. The Appendix contains some of the proofs.

[5] This approach follows recent developments in monetary theory (see, for example, Lagos and Wright (2004) and references within). If we think of the agents in our model as banks, the centralized market has a meaningful interpretation that fits well with the institutional reality of payment systems: it corresponds to the periodic settlement at the end of a specified period of time (usually a day).

[6] A novelty of our approach is that it involves non-cooperative implementation together with a Walrasian equilibrium aspect. Examples in which some form of linearity is invoked for tractability in static mechanism design environments include the classic papers by Clarke (1971) and Groves and Loeb (1975). See also Jarque (2003) for a more recent reference that deals with an intertemporal environment.

[7] Shi (1997) and Lagos and Wright (2005) study two strategic environments that eliminate distribution issues in related monetary models.

2 Preliminaries

In this section we introduce the basic economic setup and discuss some specific examples. For expository purposes, here we study the case where the size of transactions is exogenous. Several findings turn out to provide useful intuition for the general environment that we study in the next section.

Time is discrete, t, measured over the positive integers. There is a $[0, 1]$ continuum of infinitely lived agents. To generate transactions, we assume that in any given period, agents are randomly matched bilaterally. Randomness in payments is captured by assuming that an agent needs to transact with the agent he is matched with as a producer (consumer) with probability γ. Consumption gives utility u, while production gives disutility e, with $u > e$.[8] Agents have a common discount factor, $\beta \in (0, 1)$.

To keep track of production opportunities, we introduce a random variable $s \in \{0, 1\}$, which equals 1 if a meeting is a *trade meeting* (an event of probability 2γ), and 0 otherwise. Throughout the paper we will assume that production is only possible in trade meetings. In such a meeting, we let p denote the probability with which an agent chooses to produce. Thus, $p(s) \in [0, 1]$ denotes the outcome in a meeting of type s. If $s = 1$ and $p(s) = 1$, we will assume that automatically, the consumer receives utility u, and the producer receives disutility e. An *allocation* within a match is a function $p : s \to [0, 1]$. Throughout, we will study symmetric stationary allocations that can be supported by strategies that constitute perfect equilibria. We will refer to such allocations as *Incentive Feasible Allocations (IFAs).*

To familiarize the reader with the setup, let us consider the following two extremes. First, assume that agents are *anonymous* and that there are no assets in the economy. An additional important feature of this environment is that there is *no commitment.* The above assumptions rule out reputation effects, as well as any type of trade using currency or any other asset. Clearly, the only IFA then is autarky; i.e., $p(s) = 0$ for all s. Next, consider the other extreme. Assume the existence of a perfect monitoring and record-keeping

[8] The reader might wish to interpret the payment system participants as banks. In that case, we could think of each bank as being associated with a client. When a client of one bank produces for a client of another, some payment needs to be transferred to the producing party. In our setup the client-bank pair is a single economic unit. Thus, when a client produces, the client-pair bank suffers disutility, and similarly for consumption. This allows us to concentrate on the incentives within the payment system without modelling the bank-client relationship itself.

technology that allows for the types and actions of all agents to be perfectly observed and recorded in every period. In this case, a "credit" equilibrium can be sustained through a standard reputation argument. In other words, provided that β is sufficiently high, $p(s) = 1$ if $s = 1$ constitutes an IFA; i.e., a transaction takes place whenever production is desired.

The interesting cases concern situations in between these two extremes. To begin studying such cases, consider the following example. Assume that a perfect monitoring technology is not available. Instead, each agent is endowed with the ability to costlessly record his past consumption (production) with a *Payment System (PS).* Throughout the paper we will think of the PS as a planner to whom agents report their production or consumption opportunities. The PS is restricted by its technology: it cannot verify whether agents have a production opportunity within a given period. However, when it takes place, production can be verified. In that case, if, say, i produces for agent j, the balance of i with the PS is credited by "+1," while the balance of j receives an entry of "-1." We denote an agent's balance by d (an integer, not restricted in sign, no upper or lower bound). We will refer to d as the agent's *balance*. The fact that there is no bound on d can be interpreted as a PS policy that imposes no caps on individual borrowing. We now index the allocation (production decision) in a match involving an agent with balance d by $p_d(s)$.

Clearly, no agent has an incentive to build up a balance as he can always claim that he did not have a production opportunity. At the same time, since there is unlimited borrowing offered by the PS, declining to increase one's balance does not by itself decrease the probability of consuming in the future. Therefore, for any β, the only IFA is autarky. This is interesting because it suggests that caps might be necessary in order for transactions to take place. Put differently, in order for liquidity to be valuable, it has to be scarce. Next, we study the effects of caps on agents' borrowing.

We introduce caps on agents' borrowing. No penalty is imposed on agents that hit the cap other than that they cannot borrow further unless they first produce in order to improve their balance. We will demonstrate that, provided that the discount factor is sufficiently high, this policy can support a positive volume of transactions. Intuitively, the existence of a cap, C, implies that liquidity now becomes sufficiently scarce to be of value. In addition, sufficiently patient agents will produce in order to avoid being in a transaction in which the lack of liquidity prevents them from enjoying consumption. This identifies an interesting trade-off. If the PS provides little or no liquidity, by setting the cap close or equal to zero, some welfare improving transactions will not be realized. In the context of the model, this occurs if an agent faced with a consumption opportunity has hit the borrowing cap. To minimize the frequency of such inefficient instances, the PS should set the borrowing cap as high as possible. This, however, might result in the non-existence of an equilibrium with a positive volume of transactions.

We now turn to a characterization of IFAs. Let $\mathbf{p} = [p_{-C}, ..., p_0, p_1, ...]$ denote the vector of allocations, and let $\mathbf{x} = [x_{-C}, ..., x_0, x_1, ...]$ denote the distribution of agents (both in the population and per type) across states. In an IFA, associated with a positive cap, C, we have the following value functions for an agent in state d:

$$v_{-C} = \gamma(1-x_{-C})[p_{-C}(-e+\beta v_{-C+1})+(1-p_{-C})\beta v_{-C}]+[1-\gamma(1-x_{-C})]\beta v_{-C}, \quad (1)$$

$$\begin{aligned} v_{d>-C} &= \gamma \mathbf{px}[u + \beta v_{d-1}] + \gamma(1 - x_{-C})[p_d(-e + \beta v_{d+1}) + (1 - p_d)\beta v_d] \\ &\quad +[1 - \gamma px - \gamma(1 - x_{-C})]\beta v_d. \end{aligned} \quad (2)$$

The difference in the two value functions captures the fact that an agent that has hit the borrowing cap cannot consume. For the allocation to be incentive feasible, agents must be better off when they choose strategies which result in this allocation. That is, for all d such that $p_d = 1$, we must have

$$\begin{aligned} -e + \beta v_{d+1} \geq \beta v_d, \quad \forall d \geq -C, \text{ and} \\ u + \beta v_{d-1} \geq \beta v_d, \quad \forall d > -C. \end{aligned} \quad (3)$$

In addition, for all d such that $p_d = 0$, we must have

$$-e + \beta v_{d+1} \leq \beta v_d. \quad (4)$$

Discounting implies that consuming today is better than consuming at a later date. Hence, only the producer's incentive constraint is binding. In addition, note that $\mathbf{p}$ affects the law of motion of the distribution of agents across states. We concentrate on IFAs for which $x_d > 0$ for some $d > 0$. It can be shown that, under suitable parameter restrictions, the set of IFAs where $p_d = 1$, for some d, is non-empty. Autarky is always incentive feasible. The next proposition asserts that all IFAs involving a positive volume of transactions have the property that agents increase their balances up to a point. If their balance becomes sufficiently high, agents will decline opportunities to increase it further. This results in some welfare loss since production does not occur in some meetings when it would otherwise be desirable.

Proposition 1 *Assume β is sufficiently high. All stationary IFAs have the property that either $p_d = 0$ for all d, or there exists $\overline{D} \geq 0$ such that (a) $p_d = 1$, $\forall d \leq \overline{D}$, and (b) $p_d = 0$, $\forall d > \overline{D}$.*

This indicates that in order for a positive volume of transactions to take place, it must be that the cap is set sufficiently low. We now turn to the question of existence of the stationary distribution, $\mathbf{x}^*$, of agents across states. For any given distribution, we have shown that there exists $\overline{D}$ such that $p_d^* = 1$, for all $d < \overline{D}$, and $p_d^* = 0$, otherwise. Using the normalization $C = 0$ we have the following.

Proposition 2 *If β is sufficiently high, there exists a stationary IFA in which $p_d = 1$, for some d. The IFA gives rise to a uniform stationary distribution, $\mathbf{x}^*$.*[9]

Proof. Let $\overline{D}$ be such that $p_d^* = 1$, for all $d < \overline{D}$. In a stationary distribution, $\overline{D}$ does not change. The law of motion for $\mathbf{x}$ is then characterized by

$$\begin{aligned} x_0' &= x_0(1-\gamma(1-x_0)) + x_1\gamma(1-x_{\overline{D}}) \\ x_d' &= x_{d-1}\gamma(1-x_k) + x_d(1-\gamma(2-x_0-x_{\overline{D}})), \\ &\quad + x_{d+1}\gamma(1-x_0), \quad \forall 1 < d < \overline{D}, \\ x_{\overline{D}}' &= x_{\overline{D}-1}\gamma(1-x_0) + x_{\overline{D}}[1-\gamma(1-x_{\overline{D}})]. \end{aligned} \tag{5}$$

In a stationary distribution we have $x_d' = x_d$ for all d. Suppose $x_0 = x_{\overline{D}}$. The law of motion for x implies that $x_0 = x_1$ and $2x_d = x_{d+1} + x_{d-1}$. Hence, $x_d = x_{d'}$ for all $0 < d$, $d' \leq \overline{D}$ is a solution to these equations. In other words, the uniform distribution is stationary. ■

The following proposition suggests that the PS should set the cap at the maximum level consistent with the existence of an equilibrium with trade.

[9] We have not established that the stationary distribution is either unique or stable. Standard results on Markov chains cannot be directly applied as the transition matrix is state dependent.

Proposition 3 *Consider two IFAs that are supported by a uniform distribution of reserves and respective caps C and C', with $C > C'$. Welfare is higher in the allocation resulting from the greater cap, C.*

Proof. Given the agents' policy rules, and given that the distribution of money holdings is uniform, each agent will set $p_d^* = 1$ for $d < C(C')$, and $p_d^* = 0$, otherwise. In a cap y-allocation, under a uniform distribution, the probability of either consuming or producing is $y/(y+1)$. Welfare is given by $\gamma \left(y/(y+1)\right)^2$, which is clearly increasing in y. Hence, welfare is higher in the C-allocation. ■

We end this section by briefly discussing the case in which the PS can condition its policy on reports by *both* agents involved in any given transaction. More precisely, suppose that the PS can observe that i and j are in a match during a given period, and consider the following rule. If both agents report that one produced for the other, say i for j, then the PS assigns a "+1" and a "−1" to the producer and the consumer, respectively. If the consumer reports that he did *not* receive production, the producer is punished to permanent autarky, which gives a lifetime utility of 0. Assume that β is sufficiently high. Then, under the above policy, the allocation in which production takes place in each trade meeting is incentive feasible. Thus, the full information first-best allocation can be supported, and there is no need for caps. Note, however, that this scheme relies on rather strong informational requirements. For example, it is necessary that the PS can verify that i and j are in a meeting in which i is the potential producer. If this assumption is withdrawn, then i can claim (falsely) that it was j that did not produce for him, etc. In what follows, we shall restrict ourselves to environments in which the PS does not possess such information, and in which certain transactions are subject to a form of imperfect monitoring.[10]

A main feature of the models studied so far is that there is a trade-off between incentives and efficiency. In other words, the full information first-best allocation, in which the efficient quantity is produced in each trade meeting, is not incentive feasible. Caps were shown to be necessary for a positive volume of transaction to take place as part of an incentive feasible allocation. However, the existence of caps implies that the PS cannot sustain the opti-

[10] Even in such cases, the CH can accomplish more by reverting to collective punishments. For example, the CH could punish *both* i and j, say, to permanent exclusion, unless their reports were mutually consistent. We shall ignore collective punishments in what follows.

mal level of transactions in all trade meetings. The reason is reminiscent of Levine (1991): agents are unable to consume following a sequence of "lucky draws" since, in that case, they hit their cap. In addition, relaxing the cap restriction for all agents, which is accomplished by monetary injections in Levine's framework, is impossible here as this would jeopardize the existence of an equilibrium with transactions. In the next section, we study optimal PS under periodic settlement.

3 Optimal Payments under Periodic Settlement

In this section we consider the implementation of optimal transactions under private information, no commitment, and imperfect monitoring. To this end, we assume that, with probability α, the types of both agents in a meeting are observable to the PS while, with probability $1-\alpha$, the PS cannot observe whether the meeting is one in which the consumer likes what the producer can offer.[11] Both agents and the PS are assumed to know whether a transaction is monitored or not.

We have established that, if β is sufficiently high, an equilibrium can be sustained in which a transaction takes place in each verified trade meeting, under the threat that deviating agents are punished to exclusion. Can we support an allocation in which the efficient level of production takes place in *all* trade meetings, including the ones in which the opportunity to produce is private information? Henceforth, we will refer to this full information first-best allocation as the *efficient allocation.*

The difficulty lies in that with probability $(1-\alpha)$, the PS cannot verify whether a trade meeting has taken place. To see this, consider a distinguished agent who, say, for the k-th time in a row, reports that he could not produce since he had k consecutive non-trade meetings. Given the information structure, the PS can verify that the agent had k consecutive non-monitored meetings (this is an event of probability $(1-\alpha)^k$). It can also verify that the agent did not produce in any of these meetings. What the PS cannot verify, however, is whether the agent had an opportunity to produce and simply declined, or whether he did not have any trade meetings (an event of

[11] In some of what follows we will find it convenient to restrict ourselves to the case where $\alpha = 0$.

probability $(1-\gamma)^k$).

In this section we will assume that production of goods is perfectly divisible. Producing q units implies disutility $-e(q)$, while consumption of q units gives utility $u(q)$. We assume that $e'(q) > 0$, $e''(q) \geq 0$, $\lim_{q\to 0} e'(q) = 0$, and $\lim_{q\to\infty} e'(q) = \infty$. In addition, we assume $u'(q) > 0$, $u''(q) \leq 0$, $\lim_{q\to 0} u'(q) = \infty$, and $\lim_{q\to\infty} u'(q) = 0$. We will restrict ourselves to mechanisms in which the state of an agent consists of a one-dimensional real variable. More precisely, our implementation assumes that each agent has a balance, $d \in \mathbb{R}$, with the PS. For simplicity, we assume that agents' balances are perfectly observable in all transactions. The payment system prescribes how balances should be adjusted over time as a result of agents' reports about their transactions.

A number of actual transactions involve periodic *clearing* or *settlement* rounds. To study the effect of settlement in our model, we introduce a periodic pattern of length $n+1$. The first n periods of each cycle involve, as previously, bilateral transactions. This is followed by one centralized settlement round, which we model as a Walrasian market. In that market agents can trade balances for a general, non-storable good. Effectively, the Walrasian market corresponds to a settlement round in which agents that are "low" can increase their balances by producing, while those having excessive balances end up as consumers. We assume that producing ℓ units of the good gives disutility $-\ell$, while consuming ℓ units gives utility ℓ. Market clearing, thus requires that, on average, $\ell = 0$. This assumption implies that the introduction of the settlement round itself does not lead to a welfare increase. Any welfare improvements come directly from the effect of settlement on the agents' transaction patterns. The price, p, at which balances are purchased is determined by market clearing conditions.

In each of the first n periods of the cycle agents engage in bilateral transactions. In the case where a transaction is monitored, the PS observes both the type of the meeting and the quantity consumed or produced. In non-monitored meetings, however, the PS cannot observe whether an agent that is a potential producer is in a trade meeting or not. If an agent ends up producing in such meetings, however, production can be verified. On the other hand, consumption is not verifiable. Throughout, we restrict attention to outcomes that are stationary and symmetric across agents. It is useful to keep in mind the efficient allocation in this setup. We have the following.

Proposition 4 *The efficient allocation involves consumption and production* *of q^* in all bilateral trade meetings. In the settlement round we have that $E[\ell] = 0$, for all agents.*

Note that ℓ in the settlement round is indeterminate, due to linearity. However, expected (and, therefore, average) ℓ will equal zero. Clearly, if the PS can observe both agents' types; i.e., if $\alpha = 1$, the efficient allocation can be implemented provided that β is sufficiently high. In what follows we will concentrate on the case where $\alpha \in [0, 1)$. The choice of n, which refers to the frequency of the settlement rounds, is of interest, and we will discuss this further later. For now, we will simply impose that $n = 1$ and proceed to study the payment system's problem for the above setup.

3.1 The Benchmark Case: $n = 1$

In what follows, we analyze a generic period t and work backward, first considering the agent's problem in the settlement round, and then moving on to the transactions stage. Let $V(d, p)$ denote the value function of an agent that exits the transaction round with balance d. Let $v(\widehat{d}, \Psi)$ denote the value when he exits the settlement round with balance $\widehat{d}$, given that the resulting distribution of balances is given by Ψ. Taking as given the price, p, and the distribution of balances at the end of the settlement round, Ψ, agents solve the following problem at the beginning of the settlement round:

$$\begin{aligned} V(d,p) &= \max_{\ell, \widehat{d}} \{-\ell + \beta E v(\widehat{d}, \Psi)\} \\ \text{s.t.} \quad p\widehat{d} &= pd + \ell. \end{aligned} \tag{6}$$

Below we argue that, as a consequence of linearity, Ψ is degenerate. That is, all agents exit with the same balance. This feature greatly improves the tractability of the problem. In addition, it fits well with the fact that actual payment systems require centralized settlement of all debts at the end of a specified period, usually a day.

Next, we describe the problem faced by the PS during the transactions round. Recall that if an agent has an opportunity to produce in a non-monitored transaction, he can (falsely) report to the PS that he was not in a trade meeting. However, like before, if production takes place, the PS can verify the quantity produced, but consumption is not verifiable. The

PS maximizes a welfare function that involves all matches weighted by their frequency as implied by Ψ.

Let $\mathbf{d} \in \mathbb{R}^2$ denote the vector of balances of two agents that are matched during the transactions round. As mentioned earlier, we assume that balances are always observable to the PS. We shall think of agents as making reports to the PS about the type of the meeting that they are in. Agents that report a trade meeting as producers receive instructions from the PS about how much to produce. Consumers report the quantity they consumed. The PS subsequently makes balance adjustments that depend on these reports. Like before, production and delivery of goods in non-monitored transactions is verifiable, while consumption is not. In addition, agents that refuse to produce in monitored transactions are punished to permanent exclusion.

Note that, ignoring the agents' balances, there are six possibilities regarding meetings. An agent can be in a consumption, a production, or a non-trade meeting that is either monitored or non-monitored. The vector of policy rules (P_t, R_t, A_t, q_t) determines the new balances as well as the quantity produced, q_t, for agents involved in monitored transactions. These functions depend on the agents' current balances $\mathbf{d}$ and on the distribution of balances, Ψ. More precisely, $P_t(R_t)$ gives the balance adjustment for an agent who consumes(produces) in a monitored transaction, while A_t is the adjustment resulting from not trading in such a transaction. Similarly, the vector of policy rules (L_t, K_t, B_t, Q_t) determines the new balances and the quantity produced, Q_t, for agents involved in non-monitored transactions. Like before, $L_t(K_t)$ is the adjustment for an agent who consumes(produces) in a non-monitored transaction, while B_t is the adjustment for an agent who does not transact. Recall that balances are represented by real numbers not restricted in sign, while production of goods is restricted to be positive. At the end of the transactions round balances are adjusted according to reports, and agents enter the settlement round knowing their new balances.

We will concentrate on arrangements that satisfy certain incentive and participation constraints. The incentive constraints require that the following inequalities hold:

$$V(d+L,p) = V(d+B,p), \tag{7}$$
$$-e(Q)+V(d+K,p) \geq V(d+B,p). \tag{8}$$

The first incentive constraint holds with equality as consumption is not verifiable when there is no monitoring. In addition, participation constraints require that

$$
\begin{aligned}
V(d+A,p) &\geq 0, & (9)\\
V(d+B,p) &\geq 0, & (10)\\
-e(q)+V(d+R,p) &\geq 0, & (11)\\
u(q)+V(d+P,p) &\geq 0. & (12)
\end{aligned}
$$

We are now ready to formally define a payment system.

Definition 5 *A Payment System* $\mathbf{S}$ *is defined to be an array of functions* $\mathbf{S} = \{P_t, R_t, A_t, q_t; L_t, K_t, B_t, Q_t\}$. $\mathbf{S}$ *is incentive feasible if the array is chosen so as to satisfy the incentive and participation constraints.* $\mathbf{S}$ *is simple if the balance adjustments do not depend on the level of the agents' current balances. An incentive feasible* $\mathbf{S}$ *is optimal if it can support the efficient allocation.*[12]

Let $v(\mathbf{d})$ be the expected value of an agent with balance d in a meeting with an agent with balance $\widetilde{d}$. This is given by

$$
\begin{aligned}
v(\mathbf{d}) = {} & \alpha\{\gamma[u(q)+V(d+P,p)]+\gamma[-e(q)+V(d+R,p)]\\
& +(1-2\gamma)V(d+A,p)\}\\
& +(1-\alpha)\{\gamma[u(\widetilde{Q})+V(d+L,p)]+\gamma[-e(Q)+V(d+K,p)]\\
& +(1-2\gamma)V(d+B,p)]\}. \qquad (13)
\end{aligned}
$$

Recall that the efficient allocation involves production of q^* in *all* trade meetings. We have the following.

Proposition 6 *Assume that* $n=1$ *and the Friedman rule,* $p=\beta p_{+1}$, *holds. A simple incentive feasible* $\mathbf{S}$ *can support the efficient allocation in the transactions round.*

The following Proposition gives a sufficient condition in order for a simple incentive feasible payment system that supports the efficient allocation to exist.

[12] To simplify notation we shall often suppress the dependence of $\mathbf{S}$ on t.

Proposition 7 *Provided that* $(\beta - 1)p_0 d_0 > (1-\alpha)\gamma(e(q^*) - u(q^*))$*, there exists a simple optimal* $\mathbf{S}$ *that is consistent with the Friedman rule and under which agents start each new transactions round with the same balances.*

The proofs of the above two Propositions are given in the Appendix. Using the incentive and participation constraints, we can derive further properties of $\mathbf{S}$. First, note that since consumption in non-monitored transactions is not verifiable, in order for an agent that consumed to report truthfully, $\mathbf{S}$ needs to treat him the same way as if he reported a no-trade meeting; i.e., $B_t = L_t$. In addition, it must be that $p_t(K_t - B_t) \geq e(Q^*)$, $\forall t$. In other words, agents are rewarded for producing in non-monitored transactions. While the proposed PS implies that $E[\ell] = 0$ for all agents and in each settlement round, it will clearly not result in each agent having the same ℓ due to the incentive constraints.

3.2 The Case Where $n > 1$

While the $n = 1$ case is of some interest, it literally implies that settlement takes place after every transaction. In reality, periodic settlement takes place at the end of a specified period of time (usually a day). While the choice of the optimal frequency of settlement is interesting to analyze for policy purposes, here we will simply assume a settlement round after a transactions stage of length $n > 1$. In particular, we will derive sufficient conditions for a PS to implement the efficient allocation with respect to $(q^*, Q^*, E[\ell^*])$ for the general case where $n > 1$, but finite.

We distinguish between two cases: when there is discounting ($\beta < 1$) and when there is no discounting ($\beta = 1$) between the rounds of the transactions stage. However, we assume that there is never discounting between the last round of the transactions stage and the next settlement round. In addition, we assume that there is always discounting between the settlement round and the first round of the next transactions stage. In what follows, we will concentrate on history-independent PS.

Definition 8 *A PS is history-independent if the balance adjustments in any round* s*,* $1 < s \leq n$*, of the transactions stage do not depend on the history of the previous* $s - 1$ *rounds.*

We first analyze conditions under which the efficient allocation can be implemented under a Friedman rule. Then we show that, for some parameters, the Friedman rule cannot implement the efficient allocation. In contrast, the efficient allocation can always be supported by a PS that results in a constant supply of nominal balances.

3.2.1 Implementing Efficient Allocations under a Friedman Rule

Here we suggest that a PS that is consistent with the Friedman rule can support the efficient allocation. We adjust the Friedman rule to hold across n transactions rounds whenever there is positive discounting between rounds. Hence, if in period t there is a settlement round with equilibrium price p_t, then in period $t+n$ the Friedman rule requires the equilibrium price to be $p_t = \beta^n p_{t+n}$. The next Lemma pins down the value function in the settlement round, V. Due to linearity in d, it is sufficient to specify the value function for a single value of d. We do so for the equilibrium balance, $\hat{d}$.

Lemma 9 *Suppose that there is discounting $\tilde{\beta}$ between rounds in the transaction stage. Also suppose that a simple, history-independent PS follows a Friedman rule, i.e. $p_t = \beta\tilde{\beta}^{n-1} p_{t+n}$, and implements the efficient allocation. Then, the value function is given by*

$$V(\hat{d}, p) = \frac{1 - \beta\tilde{\beta}^{n-1}}{\beta\tilde{\beta}^{n-1}} p\hat{d} + \frac{\beta}{1 - \beta\tilde{\beta}^{n-1}} \left(\sum_{s=0}^{n-1} \tilde{\beta}^s \right) \gamma(u - e). \tag{14}$$

The proof of the Lemma is given in the Appendix. Consider now the following simple history-independent PS that replicates n-times the simple PS studied in the case where $n = 1$, with $X_{t+s} = \frac{X}{\beta^{n-s}}$, for all $s = 1, \ldots, n$. Here, X stands for the balance adjustment in the case where $n = 1$, and $q_{t+s} = Q_{t+s}$ are defined by

$$q_{t+s}(d, d') = \begin{cases} e^{-1}(p_t d'_t), \text{ if } d'_t \in [0, \overline{d_t}]; \\ q^*, \text{ otherwise}; \end{cases} \tag{15}$$

for all $s = 1, \ldots, n$, where $\hat{d}_t$ satisfies $e^{-1}(p_t \hat{d}_t) = q^*$. Clearly, this is an n-*times repeated* simple payment system. We have the following Proposition, whose proof can also be found in the Appendix.

Proposition 10 *Suppose that a simple payment system implements the efficient allocation when $n = 1$. Suppose that there is discounting between rounds in the transactions stage. For any $n \in N$, there exists a sufficiently high β*

such that the n-times repeated simple PS implements the efficient allocation in all n rounds. Furthermore, if there is no discounting between rounds, there exists a sufficiently high β such that replicating the simple PS n-times implements the efficient allocation for any $n \in N$.

The usual reason invoked for the optimality of the Friedman rule is that, since $\beta < 1$, producers are willing to produce less than the quantity they would produce if they could turn the proceeds into immediate consumption. In that case, the Friedman rule creates the necessary real return on balances in order to compensate for discounting. In contrast, in our environment producers are immediately rewarded for their effort in the next settlement round, and such compensation is not necessary. The Friedman rule, however, is still optimal as it gives the incentive to carry the "right" amount of balances (the one resulting in consuming q^*) into the next transactions stage. If the Friedman rule does not hold, then agents would find it beneficial to work less in the settlement round, at the cost of consuming a tiny less in the next period transactions stage.

The next result demonstrates conditions under which the efficient allocation cannot be implemented under the Friedman rule. The intuition behind the result is very similar to the one obtained in Levine (1991), but coming from a very different model.

Corollary 11 *Assume that there is discounting at rate β between the rounds of the transactions stage. Then there exists $\bar{n} \geq 1$ such that the repeated simple PS does not support the efficient allocation under the Friedman rule for any $n > \bar{n}$.*

Proof. Given the structure of the PS, all participation constraints are fulfilled if and only if the IC constraint of a non-monitored producer is fulfilled; i.e. if and only if

$$\left(\frac{1-\beta}{\beta}\right) p_t d_t + \left(\frac{\beta^n}{1-\beta^n}\right) \gamma(u-e) \geq -p_t B_t. \tag{16}$$

For any $\beta \in (0,1)$, we have that $\frac{\beta^n}{1-\beta^n} \to 0$ for $n \longrightarrow \infty$. Since the first term on the LHS is bounded and strictly less than the RHS, the inequality will be violated for n high enough. In that case the efficient allocation cannot be implemented by a Friedman rule. To complete the proof we need to demonstrate that $\left(\frac{1-\beta}{\beta}\right) p_t d_t < -p_t B$ for all β. To see that this is true, we consider the special case where $\alpha = 0$ and $n = 1$.[13] We then have

$$\begin{aligned} E[X_{t+1}] &= (\beta - 1)d_t, \\ E[X_{t+1}] &= \gamma K_{t+1} + \gamma B_{t+1}(1-2\gamma)L_{t+1}, \\ B_{t+1} &= L_{t+1}, \\ -e &= p_{t+1}\left(B_{t+1} - K_{t+1}\right). \end{aligned} \tag{17}$$

Solving the above system of equations we obtain

$$B_{t+1} + \gamma \frac{e}{p_{t+1}} = (\beta - 1)d_t.$$

Multiplying by p_t and using the fact that $\beta p_{t+1} = p_t$ we have

$$\begin{aligned} -p_t B_{t+1} &= (1-\beta)p_t d_t + \gamma e \frac{p_t}{p_{t+1}} \\ &= \beta\left[\frac{1-\beta}{\beta} p_t d_t + \gamma e\right], \end{aligned} \tag{18}$$

which implies that the desired inequality holds for any β. ■

In the next section, we show that although the Friedman rule cannot implement the efficient allocation if n is large, a policy that maintains a constant supply of nominal balances can.

3.2.2 Implementing Efficient Allocations without a Friedman Rule

We now demonstrate that the efficient allocation is implementable without relying on the Friedman rule. Indeed, the efficient allocation can be achieved through a PS in which nominal balances remain constant. In order to simplify the notation and to ease exposition, we make the following two assumptions: (i) $\alpha = 0$, i.e., there is no monitoring of any transactions, and (ii) $n = 1$. We again restrict attention to stationary payment systems, i.e., to those in which $p_t d_t$ is a constant.

Consider the following class of simple PS. Define the amount an agent can consume in a transaction by the function

$$Q_{t+1}(d, d') = \begin{cases} Q^*, & \text{if } d'_t = a; \\ 0, & \text{if } d'_t \neq a; \end{cases} \tag{19}$$

[13] Both assumptions are innocuous for two reasons. First, when $\alpha > 0$, B changes, but the argument is essentially intact. Second, we only need to check the n-th round PC, but this constraint is identical to the PC for $n = 1$.

where Q^* satisfies $u'(Q^*) = e'(Q^*)$, and $a \in \mathbb{R}$. Hence, if the balance of the potential consumer is a, the producer produces the efficient quantity, otherwise no production takes place. Note that a is an arbitrary real number and can be normalized to 0. The next Proposition asserts that the Friedman rule is not necessary to achieve the efficient allocation.

Proposition 12 *If $\beta u \geq e$, then there exists a simple payment system with $\hat{d}_t = \hat{d}$, for all t, that implements the efficient allocation.*

Proof. Again, we consider the case where $\alpha = 0$. Let $Q_t(d, d')$ be defined as in equation (19). Define the balance adjustments (K, L, B) through the following three equations

$$\begin{aligned} B &= L, \\ -e + V(d + K, p) &= V(d + B, p), \\ \gamma K + \gamma L + (1 - 2\gamma)B &= 0, \end{aligned} \tag{20}$$

where we have used the fact that Q_t is time-independent and, hence, $\hat{d}$ – as well as p – is constant over time. The first two equations express the IC constraints. The third equation ensures a law of motion on the equilibrium aggregate balances that is consistent with $\hat{d}$ being constant over time. Note that V in period t only depends on the distribution of balances through the expected costs of production in the next transactions stage, $e(Q_{t+1})$. These costs are, in turn, a function of the matching partners' choice of balances. The choice of an agent's own balance only influences the amount he can consume in the next transactions stage.

We guess that in a perfect equilibrium, given the balance adjustments (K, L, B) defined above, every agent chooses balances $\hat{d}_t = a$ in all settlement rounds. To verify that these strategies form a perfect equilibrium, we only need to verify all PCs given the definition of $Q_t(d_{t-1}, d'_{t-1})$. These are given by

$$\begin{aligned} V(\hat{d}_{t-1} + X_t, p_t) &\geq 0, \\ u + V(\hat{d}_{t-1} + B, p_t) &\geq 0, \\ -e + V(\hat{d}_{t-1} + K, p_t) &\geq 0. \end{aligned} \tag{21}$$

Since the second one is fulfilled whenever the first one holds, and the third inequality holds whenever the IC conditions hold, it is sufficient to verify that $V(\hat{d}_{t-1} + X_t, p_t) \geq 0$. Using the linearity of V, the stationarity of d and p, and the fact that expected balance adjustments are zero, we have the following.

$$\begin{aligned} V(\hat{d}_{t-1} + X_t, p_t) &= -p_t\hat{d}_t + p_t\hat{d}_{t-1} + p_tX_t + \beta E[v(\hat{d}_t, \Psi)] \\ &= p_tX_t + \beta\gamma(u-e) + \beta E[V(\hat{d}_t + X_{t+1}, p_{t+1})] \\ &= p_tX_t + \beta\gamma(u-e) + \beta V(\hat{d}_t, p_{t+1}) + \beta p_{t+1}E[X_{t+1}] \\ &= p_tX_t + \beta\gamma(u-e) + \beta V(\hat{d}_t, p_{t+1}), \end{aligned} \tag{22}$$

since $E[X_{t+1}] = 0$. Given that $\hat{d}_{t-1} = \hat{d}_t$, this yields

$$V(\hat{d}_t, p_t) = \frac{\beta}{1-\beta}\gamma(u-e). \tag{23}$$

From the definition of (K, L, B) we have that $K = (1-\gamma)\frac{e}{p}$ and $B = -\gamma\frac{e}{p}$. Hence, $V(\hat{d}_{t-1} + K_t, p_t) > V(\hat{d}_{t-1} + B_t, p_t)$ and, by the linearity of V,

$$\begin{aligned} V(\hat{d}_{t-1} + B_t, p_t) &= p_tB_t + V(\hat{d}_{t-1}, p_t) \\ &= -\gamma e + \frac{\beta}{1-\beta}\gamma(u-e) \geq 0, \end{aligned} \tag{24}$$

which holds if and only if $\beta u \geq e$. ■

Hence, we conclude that for β high enough, a simple payment system with a constant supply of balances achieves efficiency. Along the lines of the previous sections, it is straightforward to adjust the argument given here for the case where $\alpha > 0$ and for the case where $n > 1$.[14] Note that in the proof we have used, a simple PS in which the adjustment of balances depends only on the type of match that agents announce in the transactions stage, but not on individual balances. This was shown to be enough to support the efficient allocation in a perfect equilibrium. Any deviation in one's own balances does not influence the adjustment in the transactions round (i.e., $\frac{\partial K}{\partial d} = 0$). Also note that $B = L < 0$, which means that irrespective whether agents can

[14] The existence of some monitored transactions actually helps to implement the efficient outcome for lower values of β – or, equivalently, given β, for higher values of n. The reason is that $-B$ can be set lower than when $\alpha = 0$, which relaxes the (most) binding participation constraint. Also, note that the value functions V differ dependent on whether the Friedman rule is followed or not. The value functions at the end of the settlement round, $E[v(d_t, \Psi)]$ are, however, identical for the equilibrium choice of balances.

consume Q^* or are in a no-trade meeting, they are penalized with decreasing balances.

Why can the above PS implement the first-best allocation while a PS that is consistent with the Friedman rule cannot? In other words, what is the difference between the PS defined in (15) and the one defined in (19)? We summarize this as follows. If n is large enough, by reducing aggregate nominal balances, the Friedman rule is inconsistent with the implementation of q^* under all histories of transactions. A constant supply of balances, on the other hand, gets around this problem. Note that a constant supply of nominal balances implies a higher rate of inflation than the Friedman rule and, thus, might give agents the incentive to reduce their effort to accumulate balances. Indeed, under the first PS, such a policy would give agents the incentive to build lower balances than optimal, since the benefits from working less in the settlement stage would compensate the discounted cost from consuming a little less in the next transaction stage. However, under the second PS, building a slightly lower balance than the one prescribed by the PS implies a much harsher punishment (no consumption) in the next transaction round.

4 Comments

We characterized the optimal payment system in a dynamic model in which the ability of agents to perform certain welfare improving transactions is subject to random and unobservable shocks. Implementation involved assigning individual balances to agents and optimally adjusting these balances given their histories of transactions. The existence of an equilibrium in which agents transact through a payment system requires certain caps on short-term borrowing. We showed that in the absence of settlement, incentive constraints imply that the full information first-best allocation cannot be supported. The first-best is supportable if settlement is introduced, provided that it takes place with a sufficiently high frequency. If settlement rounds are infrequent, we demonstrated that the first-best allocation is not supportable within a payment system that operates under a Friedman-like rule. In contrast, a payment system policy that leads to constant supply of nominal balances implements the first-best.

A more general topic involves the study of optimal dynamic contracting in abstract private information environments in which there are periodic "full information rounds." Our model could be used to investigate several issues related to payments. In addition, given that we deal with dynamic incentives, we could investigate the time consistency of various PS policies, a problem that the current analysis abstracts from. This relates to the debate of public versus private provision of payment system services since optimal dynamic schemes might, in general, require some commitment.

References

[1] Abreu, D., Pearce, D., and E. Stacchetti (1986) "Optimal Cartel Equilibria with Imperfect Monitoring," *Journal of Economic Theory* 39, p. 251-269

[2] Abreu, D., Pearce, D., and E. Stacchetti (1990) "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica* 58 (5), p. 1041-1063

[3] Atkeson A. and R. Lucas Jr. (1993) "On Efficient Distribution with Private Information," *Review of Economic Studies*, vol. 59 (July)

[4] Cavalcanti R., Erosa A., and T. Temzelides (2004) "Liquidity, Money Creation and Destruction, and the Returns to Banking," Manuscript for the Models of Monetary Economics II, the Next Generation, *International Economic Review,* forthcoming

[5] Clarke E.H. (1971) "Multipart Pricing of Public Goods," *Public Choice* 11, p. 17-33

[6] Friedman M. (1969) "The Optimum Quantity of Money," in *The Optimum Quantity of Money and Other essays* (Aldine, Chicago, IL)

[7] Fujiki H., Green E.J., and A Yamazaki (1999) "Sharing the Risk of Settlement Failure," Federal Reserve Bank of Minneapolis Working Paper 594D

[8] Green E.J. (1987) "Lending and the Smoothing of Uninsurable Income," in *Contractual Arrangements for Intertemporal Trade,* Edward C. Prescott and Neil Wallace, Eds. Minneapolis

[9] Green, E.J., and R.H. Porter (1984) "Non-Cooperative Collusion Under Imperfect Price Information," *Econometrica* 52 (1), p. 87-100

[10] Groves T. and M. Loeb (1975) "Incentives and Public Inputs," *Journal of Public economics* 4, p. 211-226

[11] Jarque A. (2003) "Repeated Moral Hazard with Effort Persistence," Manuscript, University of Rochester

[12] Kahn C.M., and W. Roberds (1998) "Payment System Settlement and Bank Incentives," *The Review of Financial Studies* 11(4), p. 845-870

[13] Kiyotaki N., and R. Wright (1989) "On Money as a Medium of Exchange," *Journal of Political Economy* 97, p. 927-954

[14] Kiyotaki N., and R. Wright (1993) "A Search-Theoretic Approach to Monetary Economics," *American Economic Review* 83, p. 63-77

[15] Kocherlakota N. (1998) "Money is Memory," *Journal of Economic Theory* 81(2), p. 232-251

[16] Kocherlakota N. (2004) "The New Dynamic Public Finance," SED Plenary Session

[17] Lagos R. and R. Wright (2005) "A Unified Framework for Monetary Theory and Policy Analysis," *Journal of Political Economy*, forthcoming

[18] Levine D. (1991) "Asset Trading Mechanisms and Expansionary Policy," *Journal of Economic Theory* 54

[19] Mills D.C. (2003) "Alternative Central Bank Credit Policies for Liquidity Provision in a Model of Payments," Federal Reserve Board of Governors Working Paper

[20] Mirrlees J. (1971) "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies* 38, p. 175-208

[21] Myerson R.B. (1991) *Game Theory: Analysis of Conflict,* Cambridge, Mass. Harvard University Press

[22] Shi S. (1997) "A Divisible Search Model of Fiat Money," *Econometrica* 64(1), p. 75-102

[23] Spear S., and S. Srivastava (1987) "On Repeated Moral Hazard with Discounting," *Review of Economic Studies* 54(4)

[24] Temzelides T. and S. Williamson (2001) "Payments Systems Design in Deterministic and Private Information Environments," *Journal of Economic Theory* 99, p. 297-326

5 Appendix

Proof of Proposition 1: Clearly, if β becomes arbitrarily small, it is not possible to implement production even in the presence of caps. For part (a) we assume, without loss of generality, that $C = 0$. Let p^* denote a stationary allocation and x^* denote the stationary distribution of agents across states. Then

$$\begin{aligned} v_{d+1} - v_d &= \gamma \mathbf{x}^* \mathbf{p}^* \beta (v_d - v_{d-1}) + (1 - \gamma \mathbf{x}^* \mathbf{p}^*) \beta (v_{d+1} - v_d) \\ &\quad + \gamma (1 - x_0^*) p_{d+1}^* [\beta (v_{d+2} - v_{d+1}) - e] \\ &\quad - \gamma (1 - x_0^*) p_d^* [\beta (v_{d+1} - v_d) - e]. \end{aligned} \tag{25}$$

The proof of part *(a)* then follows from the following Lemma.

Lemma 13 *Suppose that $d' > d$. There is no IFA with $d \geq -C$ such that $p_d^* = 0$ and $p_{d'}^* = 1$.*

Proof. We proceed by contradiction. Without loss of generality, let $d' = d + 1$. As $p_d^* = 0$ and $p_{d+1}^* = 1$, we can use the expression for $v_{d+2} - v_{d+1}$ to get

$$\begin{aligned} \beta (v_{d+1} - v_d) &= \beta (v_{d+2} - v_{d+1}) + \frac{1 - \beta}{\gamma \mathbf{x}^* \mathbf{p}^*} (v_{d+2} - v_{d+1}) \\ &\quad + \gamma (1 - x_0^*) \beta [(v_{d+2} - v_{d+1}) - (v_{d+3} - v_{d+2})]. \end{aligned} \tag{26}$$

Since $\beta (v_{d+2} - v_{d+1}) + \frac{1-\beta}{\gamma \mathbf{x}^* \mathbf{p}^*} (v_{d+2} - v_{d+1}) > e$ and $p_d^* = 0$, it must be that $(v_{d+3} - v_{d+2}) > (v_{d+2} - v_{d+1})$. For p^* to be incentive feasible, it must then be that $p_{d+2}^* = 1$. Re-writing the expression for $v_{d+3} - v_{d+2}$, we have that $v_{d+4} - v_{d+3} > v_{d+3} - v_{d+2}$. Proceeding by induction we obtain that $v_{d+n+1} - v_{d+n} > v_{d+n} - v_{d+n-1}$, for all $n > 0$, and $p_{d+n+1}^* = 1$, for all n. Hence, for some n large enough, we must have $\beta (v_{d+n+1} - v_{d+n}) > u$, which contradicts incentive feasibility. ■

Turning to part *(b)*, it can be shown that, for β sufficiently large, in all IFAs the expression $v_{d+1} - v_d$ is strictly decreasing in d. It is easy to see that, given that $v_{d+1} - v_d$ is monotonically decreasing, we have that $v_{d+1} - v_d < \beta (v_d - v_{d-1})$. This, in turn, implies that $v_{d+1} - v_d < \beta^d (v_1 - v_0) \leq \beta^{d-1} u$, where the last inequality follows from the fact that $\beta (v_1 - v_0) < u$. Hence, there is a $\overline{D}$ sufficiently large, so that $\beta (v_{\overline{D}+1} - v_{\overline{D}}) < e$ and $p_d^* = 0$ for all $d \geq \overline{D}$.■

Proof of Proposition 6: Using the definition of v, we can rewrite the agent's problem in the settlement round as

$$\begin{aligned}
V(d,p) &= -l + pd + \max_{\underline{d}\leq \hat{d}\leq \bar{d}} -p\widehat{d} + \beta\{V(\hat{d},p_{+1}) \\
&+\alpha \int_{d'} [\gamma[u(q(\widehat{d},d')) - e(q(\widehat{d},d')) + p_{+1}(P(\widehat{d},d') + R(\widehat{d},d'))] \\
&+(1-2\gamma)p_{+1}A(\widehat{d})]d\Psi + \\
&(1-\alpha)\int_{d'} [\gamma[u(Q(\widehat{d},d')) - e(Q(\widehat{d},d')) + p_{+1}(L(\widehat{d},d') + K(\widehat{d},d'))] \\
&+(1-2\gamma)p_{+1}B(\widehat{d},d')]d\Psi\}, \qquad (27)
\end{aligned}$$

where p_{+1} denotes next period's price. Linearity of $V(d,p)$ in d implies that $V(\widehat{d},p_{+1}) = V(0,p_{+1}) + p_{+1}\widehat{d}$. Hence, we obtain the following expression.

$$\begin{aligned}
V(d,p) &= -l + pd + \beta V(0,p_{+1}) + \max_{\underline{d}\leq \hat{d}\leq \bar{d}} -p\widehat{d} + \beta p_{+1}\widehat{d} \\
&+\beta\{\alpha \int_{d'} [\gamma[u(q(\widehat{d},d')) - e(q(\widehat{d},d')) + p_{+1}(P(\widehat{d},d') + R(\widehat{d},d'))] \\
&+(1-2\gamma)p_{+1}A(\widehat{d},d')]d\Psi + \\
&(1-\alpha)\int_{d'} [\gamma[u(Q(\widehat{d},d')) - e(Q(\widehat{d},d')) + p_{+1}(L(\widehat{d},d') + K(\widehat{d},d'))] \\
&+(1-2\gamma)p_{+1}B(\widehat{d},d')]d\Psi\}. \qquad (28)
\end{aligned}$$

The first order condition with respect to $\widehat{d}$ gives

$$\begin{aligned}
&-p + \beta p_{+1} + \beta\{\alpha \int_{d'} [\gamma[u'(q(\widehat{d},d')) - e'(q(\widehat{d},d'))]\frac{\partial q}{\partial \widehat{d}} \\
&+\gamma p_{+1}[\frac{\partial (P(\widehat{d},d') + R(\widehat{d},d'))}{\partial \widehat{d}}] + (1-2\gamma)p_{+1}\frac{\partial A(\widehat{d},d')}{\partial \widehat{d}}]d\Psi \\
&(1-\alpha)\int_{d'} [\gamma[u'(Q(\widehat{d},d')) - e'(Q(\widehat{d},d'))]\frac{\partial Q}{\partial \widehat{d}} + p_{+1}[\frac{\partial (L(\widehat{d},d') + K(\widehat{d},d'))}{\partial \widehat{d}}] \\
&+(1-2\gamma)p_{+1}\frac{\partial B(\widehat{d},d')}{\partial \widehat{d}}]d\Psi\} = 0. \qquad (29)
\end{aligned}$$

Assuming that the Friedman rule holds, and considering a simple **S**, the first order condition reduces to

$$\alpha \int_{d'} \gamma[u'(q(\widehat{d}, d')) - e'(q(\widehat{d}, d'))]\frac{\partial q}{\partial \widehat{d}} d\Psi$$
$$+(1-\alpha)\int_{d'} [\gamma[u'(Q(\widehat{d}, d')) - e'(Q(\widehat{d}, d'))]\frac{\partial Q}{\partial \widehat{d}} d\Psi = 0 \tag{30}$$

Clearly, any $\widehat{d}$ that satisfies $u'(q(\widehat{d}, d')) = e'(q(\widehat{d}, d'))$ and $u'(Q(\widehat{d}, d')) = e'(Q(\widehat{d}, d'))$ also satisfies the above first order condition.■

Proof of Proposition 7: The proof proceeds by constructing an optimal payment system satisfying the degenerate distribution property. First, we demonstrate that the agents' balance decisions in the settlement round do not depend on their previous balance. In order for the agents' problem (6) to be well defined, we assume that $d \in [\underline{d}, \bar{d}]$, for some finite bound $\underline{d} < \bar{d}$. It should be clear that any stationary IF allocation with $q > 0$ and $Q > 0$ necessarily has to feature a payment system with $d \leq \bar{d}$, thus, the upper bound never binds. The lower bound requirement is akin to the familiar no Ponzi scheme condition. Later, we argue that the lower bound will also not bind for a particular choice of $[\underline{d}, \bar{d}]$.

Using the definition of v and the linearity of V we can rewrite the agent's problem in the settlement round as

$$\begin{aligned} V(d,p) &= -l + pd \\ &+ \max_{\underline{d}\leq \hat{d} \leq \bar{d}} \beta\{V(\hat{d},p) + \alpha[\gamma[u(q) - e(q) + p_{+1}(P+R)] + (1-2\gamma)p_{+1}A] \\ &\qquad + (1-\alpha)[\gamma[u(Q) - e(Q) + p_{+1}(L+K) + (1-2\gamma)p_{+1}B]]\} - p\hat{d}. \end{aligned} \tag{31}$$

The decision variable $\hat{d}$ is then independent of d due to linearity. In order to demonstrate that agents choose the same balances in the settlement round, we first construct functions $q(\hat{d}, d')$ and $Q(\hat{d}, d')$ and the corresponding PS. Let $q(d, d')$ and $Q(d, d')$ be such that $pd = e(q(d, d'))$ and $pd = e(Q(d, d'))$ for all $d < \overline{d}$ and $q(d, d') = q^*$, $Q(d, d') = Q^*$ for $d \geq \overline{d}$. We can now ensure that these lead to a unique choice of balance, $\overline{d}$, while implementing q^* and Q^*.

Next, we construct a simple PS that is consistent with the Friedman rule and show that it can support the efficient allocation in the decentralized round. Stationarity implies that $p_t\overline{d}_t = p_{t+1}\overline{d}_{t+1}$. This, together with the Friedman rule, implies that for all t, $\beta\overline{d}_t = \overline{d}_{t+1}$. The law of motion of d_{t+1} gives

$$d_{t+1} = d_t + \alpha[\gamma(P_t + R_t) + (1-2\gamma)A_t] + (1-\alpha)[\gamma(L_t + K_t) + (1-2\gamma)B_t], \quad (32)$$

which, using that $\beta\overline{d}_t = \overline{d}_{t+1}$, can be written as

$$0 > (\beta - 1)d_t = \alpha[\gamma(P_t + R_t) + (1-2\gamma)A_t] + (1-\alpha)[\gamma(L_t + K_t) + (1-2\gamma)B_t]. \quad (33)$$

In addition, the incentive and participation constraints must hold. By stationarity we have

$$(\beta - 1)p_t d_t = (\beta - 1)p_{t-1}d_{t-1} = ... = (\beta - 1)p_0 d_0. \quad (34)$$

Setting $B_t = L_t = -\frac{\gamma}{1-\gamma}u(q^*)/p_t$ and $K_t = e(q^*)/p_t$, for all t, satisfies the incentive constraints. In addition, provided that $(\beta - 1)p_0 d_0 > (1-\alpha)\gamma(e(q^*) - u(q^*))$, we can find $R_0 = P_0 = A_0 > 0$ that preserve the participation constraints and such that the following equality holds at $t = 0$

$$(\beta - 1)p_0 d_0 = (1-\alpha)\gamma(e(q^*) - u(q^*)) + \alpha p_0[\gamma(P_0 + R_0) + (1-2\gamma)A_0]. \quad (35)$$

Next, we construct $\mathbf{S}_t$ recursively. First note that the above equality pins down d_1. We can then find $R_1 = P_1 = A_1 > 0$ such that a similar equality holds at $t = 1$. This can be repeated for all $t > 1$. Thus, a sufficient condition for the existence of the desired $\mathbf{S}$ is that

$$p_0 d_0 < \frac{(1-\alpha)\gamma(u(q^*) - e(q^*))}{(1-\beta)}. \quad (36)$$

To see that the constructed PS is optimal, note that the expected balance of an agent coming from the transactions round in period t is $d_t + \alpha[\gamma(P_t + R_t) + (1-2\gamma)A_t] + (1-\alpha)[\gamma(L_t + K_t) + (1-2\gamma)B_t]$. By the law of large numbers, this equals d_{t+1}. Therefore, the expected utility of an agent in the settlement round at any date t is given by

$$\begin{array}{ll} \alpha & \{\gamma[-l - p_t(d_{t+1} - d_t - P_t)] \\ & +\gamma[-l - p_t(d_{t+1} - d_t - R_t)] \\ & +(1-2\gamma)[-l - p_t(d_{t+1} - d_t - A_t)]\} \\ +(1-\alpha) & \{\gamma[-l - p_t(d_{t+1} - d_t - L_t)] \\ & +\gamma[-l - p_t(d_{t+1} - d_t - K_t)] \\ & +(1-2\gamma)[-l - p_t(d_{t+1} - d_t - B_t)]\}, \end{array} \tag{37}$$

which can be simplified to give $-\ell$. In the transactions round, the PS implies that q^* and Q^* are produced in all trade meetings. Hence, the expected welfare under the above PS is the same as the ex-ante welfare of the efficient allocation.■

Proof of Lemma 8: For the value function at the start of the decentralized round we obtain

$$\begin{aligned} V(d_{t-n} + \sum_{s=1}^{n} X_s, p_t) &= -p_t d_t + p_t d_{t-n} + p_t \sum_{s=1}^{n} X_s + \beta E[v_1(d_t, \psi_0)] \\ &= \frac{1-\beta^n}{\beta^n} p_t d_t + p_t \sum_{s=1}^{n} X_s + (\beta + \beta^2 + \cdots + \beta^n)\gamma(u-e) \\ &\quad +\beta^n E[V(d_t + \sum_{s=1}^{n} X_s, p_{t+n})] \\ &= \frac{1-\beta^n}{\beta^n} p_t d_t + p_t \sum_{s=1}^{n} X_s + \left(\sum_{s=1}^{n} \beta^s\right) \gamma(u-e) \\ &\quad +\beta^n p_{t+n} E\left[\sum_{s=1}^{n} X_{t+s}\right] + \beta^n V(d_t, p_{t+n}) \\ &= \left(\frac{1-\beta^n}{\beta^n} - (\beta^n - 1)\right) p_t d_t + p_t \sum_{s=1}^{n} X_s \\ &\quad +\beta \left(\sum_{s=0}^{n-1} \beta^s\right) \gamma(u-e) + \beta^n V(d_t, p_{t+n}). \end{aligned} \tag{38}$$

The above expression uses that (i) $p_t = \beta^n p_{t+n}$, (ii) the payment system implements (q^*, Q^*) in each round and (iii) $E[\sum_{s=1}^{n} X_{t+s}] = (\beta^n - 1)d_t$, since the payment system implements $E[l^*] = 0$ and follows the FR. Using linearity of V and expanding V until infinity, we obtain

$$V(d,p) = \frac{1}{1-\beta^n}\left[\left(\frac{1-\beta^n}{\beta^n} - (\beta^n - 1)\right) pd + \beta\left(\sum_{s=0}^{n-1}\beta^s\right)\gamma(u-e)\right], \quad (39)$$

which yields the first result.

For the case without discounting between decentralized rounds, note that the Friedman rule and the fact that $E[l^*] = 0$ imply that $E[\sum_{s=1}^{n} X_{t+s}] = (\beta - 1)d_t$. The functional equation can then be written as

$$\begin{aligned} V(d_{t-n}, p_t) &= \left(\frac{1-\beta}{\beta} - (\beta - 1)\right) pd + \beta n\gamma(u-e) + \beta V(d_t, p_{t+n}) \\ &= \frac{1}{1-\beta}\left[\left(\frac{1-\beta}{\beta} - (\beta - 1)\right) pd + \beta n\gamma(u-e)\right], \end{aligned} \quad (40)$$

which yields the second result.■

Proof of Proposition 9: First, assume that there is discounting between transactions rounds. Under the PS we consider, the FONC has the same form as for the case where $n = 1$. This is because the functions (q_s, Q_s) only depend on initial balances $d_{t-n,}$ and the PS is simple and history-independent. Hence, the Friedman rule, together with the functions $q(\cdot)$ and $Q(\cdot)$, result in a unique choice of $\hat{d}_t$ in every settlement round. This implies that the distribution of balances, Ψ, after the settlement round is degenerate and depends only on $\hat{d}_t$ and on the PS itself. Using that, the stochastic process of transactions is *iid* over time and assuming that a Law of Large Numbers holds across agents, expected balances have to decrease in accordance with the Friedman rule. This implies that

$$E\left[\sum_{s=1}^{n} X_{t+s}\right] = (\beta^n - 1)\, d_t. \quad (41)$$

Next, we verify that the IC and PC hold in each period. In particular, note that for the n-th round IC constraints we have that $\{K_t, L_t, B_t\}$ are independent of the history of balance adjustments, $\sum_{s=1}^{n-1} X_{t-n+s}$. Hence, the n-th round IC constraints are identical to the ones for the case where $n = 1$.

Consider now the IC for an agent in a non-monitored transaction in round $n-1$. In that case, for all $\sum_{s=1}^{n-2} X_{t-n+s}$, we have

$$\begin{aligned}
& -e+\beta E\left[v_n(d_{t-n}+\sum_{s=1}^{n-2} X_{t-n+s}+K_{t-1},\Psi_{t-1})\right] \\
\geq \; & \beta E\left[v_n(d_{t-n}+\sum_{s=1}^{n-2} X_{t-n+s}+B_{t-1},\Psi_{t-1})\right], \text{ or} \\
& -e+\beta\gamma(u-e)+\beta E\left[V(d_{t-n}+\sum_{s=1}^{n-2} X_{t-n+s}+K_{t-1}+X_t,p_t)\right] \\
\geq \; & \beta\gamma(u-e)+\beta E\left[V(d_{t-n}+\sum_{s=1}^{n-2} X_{t-n+s}+B_{t-1}+X_t,p_t)\right], \text{ or} \\
& -e+\beta V(d_{t-n},p_t)+\beta p_t(K_{t-1}+E[X_t]) \\
\geq \; & \beta V(d_{t-n},p_t)+\beta p_t(B_{t-1}+E[X_t]), \text{ or} \\
& -e+\beta p_t K_{t-1}\geq \beta p_t B_{t-1}, \qquad (42)
\end{aligned}$$

where we have used linearity of V in real balances. One can iterate over s to obtain IC constraints of the form

$$-e+\beta^{n-s}p_t K_{t-n+s}\geq \beta^{n-s}p_t B_{t-n+s}, \qquad (43)$$

for all $s=1,\ldots,n$. Using the definition of X_{t-n+s}, for all s, all IC constraints are fulfilled whenever those of the simple PS are fulfilled.

Coming to the PC, assume without loss of generality that $X_t > B_t$, for all X_t; i.e., that B_t is, in each period, the worst possible adjustment in terms of incentives to participate. Clearly, receiving B forever is the worst possible path for balance adjustments within the n rounds. Thus, for the n-th round we obtain

$$
\begin{aligned}
& V(d_{t-n} + \sum_{s=1}^{n-1} X_s + B_n, p_t) \\
= & \frac{1-\beta^n}{\beta^n} p_t d_t + \frac{\beta}{1-\beta} \left(\sum_{s=0}^{n-1} \beta^s \right) \gamma(u-e) + p_t \sum_{s=1}^{n} B_{t-n+s} \\
= & \left(\frac{1-\beta^n}{\beta^n} \right) p_t d_t + \left(\frac{\sum_{s=1}^{n} \beta^s}{1-\beta^n} \right) \gamma(u-e) \geq -p_t B \left(1 + \frac{1}{\beta} + \cdots + \frac{1}{\beta^{n-1}} \right) \\
= & \left(\frac{1-\beta^n}{\beta^n} \right) p_t d_t + \left(\frac{\sum_{s=1}^{n} \beta^s}{1-\beta^n} \right) \gamma(u-e) \geq -p_t B \left(\frac{\sum_{s=1}^{n} \beta^s}{\beta^n} \right) \\
= & \left(\frac{1-\beta}{\beta} \right) p_t d_t + \left(\frac{\beta^n}{1-\beta^n} \right) \gamma(u-e) \geq -p_t B. \qquad (44)
\end{aligned}
$$

Clearly, given n, this inequality is satisfied for β close to one since the second term in the LHS grows unboundedly large, while the RHS remains constant. Note that all other participations constraints in the n-th round hold whenever the PC for the worst possible adjustment of balances is satisfied. Finally, all s-th round participation constraints ($s < n$) are fulfilled if the n-th round constraints hold since

$$
\begin{aligned}
& \beta E \left[v_n(d_{t-n} + \sum_{s=1}^{n-2} X_{t-n+s} + B_{t-1}, \Psi_{t-1}) \right] \\
& \quad = \beta\gamma(u-e) + \beta E \left[V(d_{t-n} + \sum_{s=1}^{n-2} X_{t-n+s} + B_{t-1} + X_t, p_t) \right] \\
= & \ \beta\gamma(u-e) + \beta V(d_{t-n}, p_t) + \beta p_t \sum_{s=1}^{n-2} X_{t-n+s} + B_{t-1} + \beta p_t E[X_t] \geq 0 \qquad (45)
\end{aligned}
$$

The last inequality follows from the n-th round PC. This completes the proof of the first statement.

For the case where there is no discounting between rounds, we consider the PS given by $X_{t+s} = X$, for all $s = 1, \ldots, n$, with $q_{t+s}(d, d')$ and $Q_{t+s}(d, d')$ defined exactly as for the case where $n = 1$. The FONC has the same structure as in the previous case. Thus, in the case where the law of motion for nominal balances is consistent with the Friedman rule, equilibrium real balances are, again, independent of Ψ. The adjustment of balances is given by

$$E\left[\sum_{s=1}^{n} X_{t+s}\right] = (\beta-1)d_t, \tag{46}$$

which ensures that the law of motion follows the Friedman rule and that $E[l^*] = 0$. As in the previous case, one can then verify that the incentive constraints are fulfilled in every round of the transactions stage. To check the participation constraints in the n-th round, we again identify B with the worst possible adjustment. Then,

$$V(d_{t-n}+\sum_{s=1}^{n-1} X_{t-n+s}+B_t,p_t) = \frac{1-\beta}{\beta}p_t d_t+\frac{\beta}{1-\beta}n\gamma(u-e)+np_t B_t \geq 0. \tag{47}$$

For $n = 1$, and β sufficiently close to one, we obtain

$$\frac{\beta}{1-\beta}\gamma(u-e) \geq p_t B_t. \tag{48}$$

Thus, the above inequality is also satisfied for any $n \in N$. As in the first part of the proof, if the participation constraints hold for the n-th round, they also hold in all previous rounds. This completes the proof of the second statement.■