Transactions, Settlement, 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

First version: July 2004

September 17, 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. The third author gratefully acknowledges support from the NSF through grant SES 0517862.

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 implemented. We introduce a periodic pattern in which several rounds of bilateral transactions are followed by centralized settlement. We demonstrate that the first-best is implementable, provided that settlement takes place with a sufficiently high frequency. The optimal adjustments on participants' balances by the payment system need not follow a Friedman 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 dynamic 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 the participants' abilities 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 in order 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

[1] To get some idea of the magnitudes involved, say, 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 billion, with a daily average of about €1.7 billion. 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 million.

[2] See Kahn and Roberds (1998) for a well-cited paper in this literature. Temzelides and Williamson (2001) investigate some of the issues studied here. Their model and conclusions, however, are very different from ours.

(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 natural 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 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. Finally, this setup naturally lends itself to mechanism design. Unlike the standard monetary theory approach, however, our model involves *no currency*.

Our approach is also related to the dynamic contracting literature.[3] Agents in our model participate in two kinds of transactions. Some transactions can be perfectly monitored, while other transactions are subject to imperfect monitoring. In the latter case, the payment system needs to induce truthful revelation. 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]

We find that a positive volume of transactions requires that 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 welfare loss, 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 since a positive fraction of agents reach the cap in any given period.

Actual payments systems involve the periodic *settlement* of debts at the end of a specified period of time (usually a day or a month). We proceed

[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 desired transaction must be incentive compatible for both participants.

by imposing 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 case where there is none. In order to simplify the contracting environment, we assume linear utility in the settlement stage.[5] 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. As is well known, 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 implementing the first best *independent* of the distribution.[6]

We find that the first-best is implementable if settlement takes place with a sufficiently high frequency. The significance of this result lies in the fact that the introduction of periodic settlement rounds in our model does not increase welfare by itself. The welfare increase is accomplished indirectly, through the interplay between settlement and intertemporal incentives. Interestingly, while payment systems that operate under a version of the Friedman rule can be optimal, this property is not necessary for efficiency. This somewhat surprising conclusion results from the fact that the payment system in our model can tailor rewards and punishments to individual participants. This is akin to individual-specific tax rates that could be made dependent on an agent's history of reports. Kocherlakota (2005) argues informally that the Friedman rule might not be needed for efficiency if the planner can impose individual-specific taxes. Our model provides implement for this observation.

The paper proceeds as follows. Section 2 introduces a simple version of the basic economic environment subject to indivisibility restrictions which we

[5] 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. See Lagos and Wright (2004) for an example of a monetary model in which trade is periodically centralized.

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

subsequently drop. The main finding is that the first-best allocation is not implementable in the absence of settlement. This finding carries over to the general case studied later. The main topic of the paper, concerning optimal settlement under private information, is studied in Section 3. There we demonstrate that the first-best can be implemented, provided that settlement is sufficiently frequent. A brief conclusion follows. The Appendix contains some of the proofs.

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 $u > e$.[7] Agents have a common discount factor, $\beta \in (0, 1)$.

To keep track of production opportunities, we find in convenient to 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 implemented by strategies that constitute perfect

[7] 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.

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,* there is *no commitment,* and there are no assets in the economy. Clearly, these assumptions rule out reputation effects, as well as any type of trade using currency or any other asset. Thus, the only IFA is autarky; i.e., $p(s) = 0$ for all s. Next, consider the other extreme. Assume the existence of a monitoring and record-keeping 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. For an example, assume that a monitoring technology is not available but each agent is endowed with the ability to costlessly record his past consumption (production) with an entity which we will call the *Payment System (PS).* Throughout the paper we will think of the PS as a central planner to whom agents report their production or consumption opportunities. The PS 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). 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, in the absence of a cap on balances, 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, unlimited borrowing implies that 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 in this case is autarky. This suggests that caps might be necessary in order for transactions to take place. Put differently, in order for liquidity to be valuable, it must be scarce.[8]

[8] The reader might wonder about the case in which the PS can condition its policy on reports by *both* agents involved in a transaction. Provided that β is sufficiently high, it can be shown that the first-best allocation can be implemented in that case, and there is no need for caps. This, however, relies on rather strong informational requirements. For

Next, we study the effects of introducing 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 implement a positive volume of transactions. Intuitively, the existence of a cap 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:

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

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

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

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

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

$$-e + \beta v_{d+1} \leq \beta v_d. \tag{4}$$

Discounting implies that consuming today is better than consuming at a later date. Hence, only the producer's incentive constraint is binding. 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 that $C = 0$ 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, $\mathbf{x}^*$.*

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.

Proposition 3 *Consider two IFAs that are implemented 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. ■

To summarize, the main feature of the model studied so far is that it involves 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. The reason for this is simple. Caps were shown to be necessary for a positive volume of transaction to take place as part of an incentive feasible allocation. However, the very existence of caps implies that the PS cannot sustain the first best level of transactions, since agents hit the cap with positive probability. This reasoning 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 a positive volume of transactions. While in this section we studied a model that involves indivisibilities, this principle holds true more generally. 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 volume of transactions under private information, no commitment, and imperfect monitoring.

We will extend the model of the previous section in several directions. First, in order to capture the fact that transactions are subject to imperfect monitoring, 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.[9] Both agents and the PS are assumed to know whether a transaction is monitored or not. While the opportunity to produce may not be verifiable, production itself, when it takes place, is always verified. We will assume, however, that consumption is not verifiable.

As in the previous section, we can establish that, if β is sufficiently high, an equilibrium can be sustained in which a transaction takes place in each verified trade meeting, This allocation is supported under the threat that deviating agents are punished to exclusion. Can we implement 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 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. For example, 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 probability $(1-\gamma)^k$).

In this section we will further generalize the model by assuming 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 $\lim_{q\to\infty} u'(q) = 0$. Thus, there exists a unique *optimal* q^*such that $u^{'}(q^*) = e^{'}(q^*)$. In order to simplify notation, in what follows we will denote $u(q^*)$ by u and $e(q^*)$ by e. We will restrict ourselves to mechanisms in which the state of an agent consists of a

[9] In some of what follows we will restrict ourselves to the special case where $\alpha = 0$. This is the case most commonly studied in the monetary theory literature.

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 PS prescribes how balances should be adjusted over time as a result of agents' reports about their transactions. A moment's reflection should convince the reader that as in the previous section, the following Proposition holds.

Proposition 4 *Provided that $\alpha \in [0,1)$, the allocation in which q^* is produced in every trade meeting is not implementable.*

The intuition behind this result is simple. If an agent always receives q^* independent of his history of reports, there is no incentive for him to ever produce during non-monitored meetings. In that case, production does not occur in such meetings. Thus, an agent must receive a quantity less than q^* in at least some occasions.

Can this difficulty be overcome? We will demonstrate that a small amendment to our basic model can overcome the difficulty resulting from imperfect monitoring. Furthermore, the amended model involves a feature that corresponds well with the institutional reality of actual payments systems, namely 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. During settlement, agents can trade balances for a general, non-storable good. Effectively, during settlement 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 (aggregate), $\ell = 0$. Importantly, this 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. Recall that 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 transactions, however, the PS cannot observe whether an agent that is a potential producer is in a trade meeting or not. Also, it is important to recall that if an agent ends up producing in such meetings production can be verified while, 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. It is easy to verify the following.

Proposition 5 *The (ex ante) efficient allocation involves consumption and production of q^* in all bilateral trade meetings. Efficiency implies that, prior to the settlement round, $E[\ell] = 0$, for all agents.*

Note that ℓ is indeterminate, due to linearity. However, expected (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 optimal payment system given 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:

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

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 settlement of all debts at the end of a specified period, usually a day or a month.

Next, we describe the problem faced by the PS during the transactions round. 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, we assume that 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.

Observe 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 in general may 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:

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

The first incentive constraint holds with equality as consumption is not verifiable when there is no monitoring. In addition, participation constraints require that agents are better off staying in the system, i.e.,

$$V(d+A,p) \geq 0, \quad (9)$$
$$V(d+B,p) \geq 0, \quad (10)$$
$$-e(q)+V(d+R,p) \geq 0, \quad (11)$$
$$u(q)+V(d+P,p) \geq 0. \quad (12)$$

We are now ready to formally define a payment system.

Definition 6 *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 satisfies the incentive and participation constraints.* $\mathbf{S}$ *is simple if balance adjustments do not depend on the agents' current balances. An incentive feasible* $\mathbf{S}$ *is optimal if it can implement the efficient allocation.*[10]

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

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

The next Proposition presents a necessary and sufficient condition for an optimal PS to exist. This condition is that $\beta u \geq e$. Interestingly, this inequality is identical to the one required for the existence of monetary equilibria in related monetary models. It can be shown that, provided that the

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

above condition holds, a variety of PS can be used to implement the optimal allocation. Examples include PS that operate under a Friedman rule, under which nominal balances shrink at the rate of time preference, as well as PS in which nominal balances remain constant over time. In what follows we will concentrate on simple PS.

Proposition 7 *There exists a simple optimal PS such that $\hat{d}_t = \hat{d}$, for all t, if and only if $\beta u \geq e$.*

Proof. For simplicity, we consider the case where $\alpha = 0$. Let $Q_t(d, d')$ be defined by

$$Q_{t+1}(d, d') = \begin{cases} Q^*, & \text{if } d'_t = a; \\ 0, & \text{if } d'_t \neq a; \end{cases} \tag{14}$$

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. Define the balance adjustments (K, L, B) through the following three equations

$$B = L, \tag{15}$$

$$-e + V(d + K, p) = V(d + B, p), \tag{16}$$

$$\gamma K + \gamma L + (1 - 2\gamma) B = 0, \tag{17}$$

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. Let X_t denote the expected balance adjustment prior to the period t transaction round. To verify that these strategies form a perfect equilibrium, we only need to verify that all PCs hold. These are given by

$$\begin{aligned} V(\hat{d}_{t-1} + X_t, p_t) &\geq 0, && (18) \\ u + V(\hat{d}_{t-1} + B, p_t) &\geq 0, && (19) \\ -e + V(\hat{d}_{t-1} + K, p_t) &\geq 0. && (20) \end{aligned}$$

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 $E[X_{t+1}] = 0$, we have the following.

$$\begin{aligned} V(\hat{d}_{t-1} + X_t, p_t) &= -p_t\hat{d}_t + p_t\hat{d}_{t-1} + p_tX_t + \beta E[v(\hat{d}_t, \Psi)] \\ &= p_tX_t + \beta\gamma(u - e) + \beta E[V(\hat{d}_t + X_{t+1}, p_{t+1})] \\ &= p_tX_t + \beta\gamma(u - e) + \beta V(\hat{d}_t, p_{t+1}) + \beta p_{t+1}E[X_{t+1}] \\ &= p_tX_t + \beta\gamma(u - e) + \beta V(\hat{d}_t, p_{t+1}). && (21) \end{aligned}$$

Given that $\hat{d}_{t-1} = \hat{d}_t$, this yields

$$V(\hat{d}_t, p_t) = \frac{\beta}{1 - \beta}\gamma(u - e). \tag{22}$$

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,

$$\begin{aligned} V(\hat{d}_{t-1} + B_t, p_t) &= p_tB_t + V(\hat{d}_{t-1}, p_t) \\ &= -\gamma e + \frac{\beta}{1 - \beta}\gamma(u - e) \geq 0, && (23) \end{aligned}$$

which holds if $\beta u \geq e$.

For the converse, suppose that $\beta u < e$. Let ρ_{t+1} denote the aggregate change in balances in period $t + 1$. Since $B_{t+1} = L_{t+1}$, the following three conditions must hold for a simple PS to be optimal.

$$\begin{aligned} &[\gamma K_{t+1} + (1 - \gamma)B_{t+1}] = E[X_{t+1}] = (\rho_{t+1} - 1)d_t && (24) \\ &p_{t+1}K_{t+1} - e \geq p_{t+1}B_{t+1} && (25) \\ &\frac{1 - \rho_{t+1}}{\rho_{t+1}}p_{t+1}d_{t+1} + \frac{\beta}{(1 - \beta)}\gamma(u - e) \geq -p_{t+1}B_{t+1}. && (26) \end{aligned}$$

Note that the last inequality is the participation constraint, $V(d_t+Bt+1, p_{t+1}) \geq 0$, where we have used the fact that $p_{t+1}(d_t - d_{t+1}) = \frac{1-\rho_{t+1}}{\rho_{t+1}} p_{t+1} d_{t+1}$. We can replace $p_{t+1}K_{t+1}$ everywhere to obtain two inequalities that involve only B_{t+1}:

$$\frac{\rho_{t+1}-1}{\rho_{t+1}} p_{t+1} d_{t+1} - \gamma e \geq p_{t+1} B_{t+1}, \text{ and} \tag{27}$$

$$\frac{1-\rho_{t+1}}{\rho_{t+1}} p_{t+1} d_{t+1} + \frac{\beta}{(1-\beta)} \gamma (u-e) \geq -p_{t+1} B_{t+1}. \tag{28}$$

Combining these two we obtain

$$\frac{1-\rho_{t+1}}{\rho_{t+1}} p_{t+1} d_{t+1} + \frac{\beta}{(1-\beta)} \gamma (u-e) \geq \frac{1-\rho_{t+1}}{\rho_{t+1}} p_{t+1} d_{t+1} + \gamma e, \tag{29}$$

or

$$\frac{\beta}{(1-\beta)} \gamma (u-e) \geq \gamma e. \tag{30}$$

This is a contradiction. ■

A few remarks are in order. First, although the PS we constructed in the above proof operates under constant nominal balances, it can be shown that, provided that $\beta u \geq e$, a variety of PS can be used to implement the optimal allocation. For example, as mentioned earlier, a PS that operates under a Friedman rule, under which nominal balances shrink at the rate of time preference, is also optimal. It is also straightforward to adjust the argument given here for the case where $\alpha > 0$ and for the case where $n > 1$.[11]

Using the incentive and participation constraints, we can point to further properties of **S**. First, note that since consumption in non-monitored transactions is not verifiable, in order for an agent that consumed to report truthfully, **S** needs to treat him the same way as if he reported a no-trade meeting; i.e., $B = L < 0$, which means that irrespective whether agents can consume Q^* or are in a no-trade meeting, they are penalized with decreasing balances. In addition, it must be that $p_t(K-B) \geq e(Q^*)$, $\forall t$. In other words, agents are rewarded for producing in non-monitored transactions. Finally, it should be clear that while the proposed PS implies that $E[\ell] = 0$ for all agents, it will clearly not result in each agent having the same ℓ due to the incentive constraints.

[11] The existence of some monitored transactions actually helps to implement the efficient outcome for lower values of β – or, as we will see later, 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.

3.2 Settlement Frequency

The case where $n = 1$ is of limited interest as 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 or a month). 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 an optimal allocation with respect to $(q^*, Q^*, E[\ell^*])$ for the general case where $n > 1$.

We will assume that agents discount factor during each transaction round is $\tilde{\beta}$. We will consider two cases: when there is discounting ($\tilde{\beta} = \beta$) and when there is no discounting ($\tilde{\beta} = 1$). However, we assume that there is no 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. We have the following.

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 transactions in the previous $s - 1$ rounds.*

For simplicity, in what follows, we will concentrate on history-independent, simple PS.

The next result demonstrates conditions under which an optimal simple repeated PS does not exist. The intuition behind the result is similar to the one obtained in Levine (1991), but coming from a very different model.

Proposition 9 *Assume that $\tilde{\beta} = \beta < 1$. Consider the repeated PS that replicates n-times an optimal simple PS for the case where $n = 1$. This PS is optimal if and only if $\beta^n u > e$. Thus, there exists $\bar{n} \geq 1$ such that a repeated simple PS is optimal if and only if $n < \bar{n}$.*

Proof. Denote the aggregate change in balances in period $t+1$ by ρ_{t+1} and let $\rho = \Pi_1^n \rho_{t+s}$ be the total aggregate change in balances across the n-period cycle. Assuming a simple repeated PS, balances within the cycle that are given by

$$X_{t+s} = \frac{X_{t+n}}{\beta^{n-s}} \tag{31}$$

for all $s = 1, \ldots, n$, where $X_{t+n} = X$ is the balance adjustment for the optimal PS when $n = 1$.

First, recall that for a simple repeated PS, whenever the PC for the n-th period and for $\sum_{s=1}^{n} B_{t+s}$ are fulfilled, all other PCs are also fulfilled. Hence,

$$\frac{1-\rho}{\rho} p_{t+n} d_{t+n} + \frac{\beta}{1-\beta^n} \left(\sum_{s=0}^{n-1} \beta^s \right) \gamma(u-e) \geq -p_{t+n} \sum_{s=1}^{n} B_{t+s}. \tag{32}$$

Second, using $B_{t+s} = L_{t+s}$, for all $s = 1, \ldots, n$, the aggregate law of motion is given by

$$\left[\gamma \sum_{s=1}^{n} K_{t+s} + (1-\gamma) \sum_{s=1}^{n} B_{t+s} \right] = E[X_{t+n}] = (\rho - 1) d_t. \tag{33}$$

Third, in every period, the IC is given by

$$-e + \beta^{n-s} p_{t+n} K_{t+s} \geq \beta^{n-s} p_{t+n} B_{t+s}, \tag{34}$$

for all $s = 1, \ldots, n$ or. Using the definition of X_{t+s}, this can be written as

$$-e + p_{t+n} K_{t+n} \geq p_{t+n} B_{t+n}. \tag{35}$$

Note that the last equation implies that the IC are *identical for all periods of the cycle*. This property follows from the construction of the simple repeated PS. Now, using the fact that

$$\sum_{s=1}^{n} X_{t+s} = \frac{1-\beta^n}{\beta^{n-1}(1-\beta)} X_{t+n}, \tag{36}$$

we obtain for the law of motion that

$$p_{t+n} K_{t+n} = \frac{1}{\gamma} \frac{\beta^{n-1}(1-\beta)}{1-\beta^n} \frac{\rho - 1}{\rho} p_{t+n} d_{t+n} - \frac{1-\gamma}{\gamma} p_{t+n} B_{t+n}. \tag{37}$$

Replacing $p_{t+n} K_{t+n}$ in the IC we obtain two inequalities that involve only $\sum_{s=1}^{n} B_{t+s}$:

$$\frac{\rho - 1}{\rho} p_{t+n} d_{t+n} - \gamma e \frac{1-\beta^n}{\beta^{n-1}(1-\beta)} \geq p_{t+n} \sum_{s=1}^{n} B_{t+s}, \text{ and} \tag{38}$$

$$\frac{1-\rho}{\rho} p_{t+n} d_{t+n} + \frac{\beta}{(1-\beta)} \gamma(u-e) \geq -p_{t+n} \sum_{s=1}^{n} B_{t+s}. \tag{39}$$

The efficient allocation is implementable if and only if both inequalities are fulfilled. This is the case whenever

$$\frac{1-\rho}{\rho}p_{t+n}d_{t+n} + \frac{\beta}{1-\beta}\gamma(u-e) \geq \frac{1-\rho}{\rho}p_{t+n}d_{t+n} + \gamma e\frac{1-\beta^n}{\beta^{n-1}(1-\beta)}, \quad (40)$$

or, whenever

$$\beta^n u \geq e. \quad (41)$$

Finally, note that, if n is large, the LHS of the above condition converges to zero. Thus, for every $\beta < 1$, there exists N such that for $n > N$ an optimal simple repeated PS does not exist. ■

The following follows as a corollary of the above Proposition.

Proposition 10 *Assume that $\tilde{\beta} = 1$. For any $n \in N$, an optimal repeated simple PS exists if and only if $\beta u > e$, i.e., if there exists a simple optimal PS when $n = 1$. The optimal PS in this case replicates n-times the optimal PS for the case where $n = 1$.*

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 implemented. The first-best is implementable if settlement is introduced, provided that it takes place with a sufficiently high frequency. In contrast, if settlement rounds are infrequent, we demonstrated that the first-best allocation is not implementable, at least within a restricted class of payment systems.

Our conclusions about the welfare improving role of settlement fit well with the institutional reality of actual payment systems. Indeed, settlement has been an integral part of both whole-sale net systems, in which it occurs at the end of each day, as well as of retail payment systems, (e.g., credit cards), in which settlement occurs usually at the end of each month. On the other hand, our current model cannot be used to study the question of optimal settlement frequency since we abstract from any costs associated with increased frequency. In that regard, by introducing a dynamic general equilibrium model in which incentive constraints rationalize a welfare improving role for settlement we have only scratched the surface.

Our model could be used to investigate several other issues related to payments. An open question is whether more complicated payment systems than the ones considered in this analysis could implement the first-best under less restrictive conditions. Although we have not studied this issue formally, we suspect that the answer is no, in general. In addition, one could extend our analysis to study properties of reasonable payment systems in the case when the first-best cannot be supported. Finally, 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. A more general topic involves the study of optimal dynamic contracting in abstract private information environments in which there are periodic "full information rounds."

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] Kocherlakota N. (2005) "Optimal Monetary Policy: What we know and What we Don't Know" *International Economic Review* 46(2) p. 715-729

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

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

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

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

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

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

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

[25] 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

$$\begin{aligned} v_{d+1} - v_d &= \gamma \mathbf{x}^* \mathbf{p}^* \beta(v_d - v_{d-1}) + (1 - \gamma \mathbf{x}^* \mathbf{p}^*)\beta(v_{d+1} - v_d) \\ &\quad + \gamma(1 - x_0^*) p_{d+1}^* [\beta(v_{d+2} - v_{d+1}) - e] \\ &\quad - \gamma(1 - x_0^*) p_d^* [\beta(v_{d+1} - v_d) - e]. \end{aligned} \tag{42}$$

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

Lemma 11 *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 obtain

$$\begin{aligned} \beta(v_{d+1} - v_d) &= \beta(v_{d+2} - v_{d+1}) + \frac{1-\beta}{\gamma \mathbf{x}^* \mathbf{p}^*}(v_{d+2} - v_{d+1}) \\ &\quad + \gamma(1 - x_0^*)\beta[(v_{d+2} - v_{d+1}) - (v_{d+3} - v_{d+2})]. \end{aligned} \tag{43}$$

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 $d \geq \overline{D}$.■